



Többcímkes mélytanuló hálózatok optimalizálása orvosi képfeldolgozási feladatokra

Egyetemi doktori (PhD) értekezés

A szerző neve: Katona Tamás
Témavezető neve: Dr. Harangi Balázs

DEBRECENI EGYETEM
Természettudományi és Informatikai Doktori Tanács
Informatikai Tudományok Doktori Iskola
Debrecen, 2024.

Ezen értekezést a Debreceni Egyetem Természettudományi és Informatikai Doktori Tanács Informatikai Tudományok Doktori Iskola Diszkrét matematika, adatfeldolgozás és vizualizáció programja keretében készítettem a Debreceni Egyetem műszaki doktori (PhD) fokozatának elnyerése céljából.

Nyilatkozom arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét.

Debrecen,

.....
a jelölt aláírása

Tanúsítom, hogy Katona Tamás doktorjelölt 2022-2024. között a fent megnevezett Doktori Iskola Diszkrét matematika, adatfeldolgozás és vizualizáció programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult.

Nyilatkozom továbbá arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét.

Az értekezés elfogadását javasolom.

Debrecen,

.....
a témavezető aláírása

Többcímkes mélytanuló hálózatok optimalizálása orvosi képfeldolgozási feladatokra

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
a informatikai tudományágban

Írta: Katona Tamás okleveles programtervező informatikus

Készült a Debreceni Egyetem Informatikai Tudományok doktori
iskolája
Diszkrét matematika, adatfeldolgozás és vizualizáció programja
keretében.

Témavezető: Dr. Harangi Balázs

Az értekezés bírálói:

Dr.

Dr.

Dr.

A bírálóbizottság:

elnök:

Dr.

tagok:

Dr.

Dr.

Dr.

Dr.

Az értekezés védésének időpontja:

Tartalomjegyzék

| | |
|----------------------------------------------------|-----------|
| Nyelvhasználat | IX |
| Rövidítések jegyzéke | X |
| Szójegyzék | XI |
| 1. Bevezetés | 1 |
| 2. Alapkonceptiók és módszerek | 7 |
| 2.1. Használt adatok | 7 |
| 2.1.1. ChestX-ray14 | 8 |
| 2.1.2. MIMIC-CXR | 8 |
| 2.1.3. Adatok összehasonlítása | 9 |
| 2.1.4. Címkék reprezentálása | 11 |
| 2.1.5. Adatok orvosi háttere | 11 |
| 2.2. Definíciók | 16 |
| 2.2.1. Digitális kép | 16 |
| 2.2.2. Képaugmentáció | 17 |
| 2.2.3. Bináris osztályozás | 18 |
| 2.2.4. Többcímkes osztályozás | 19 |
| 2.2.5. Metrikák | 20 |
| 2.2.6. Aktivációs és veszteségfüggvények | 23 |
| 2.3. Algoritmusok | 26 |
| 2.3.1. SGD | 26 |
| 2.3.2. Adam | 27 |
| 2.3.3. RMSProp | 28 |
| 2.3.4. Keresztvalidáció | 29 |
| 3. Többcímkes osztályozó építése | 31 |
| 3.1. Konvolúciós neurális hálózatok | 32 |
| 3.2. Regularizációs technikák | 34 |
| 3.3. Javasolt architektúra | 36 |
| 3.4. Eredmények | 37 |

| | |
|---------------------------------------------------------------------------------------|------------|
| 3.5. Összefoglalás | 40 |
| 4. Hiperparaméter-optimalizált osztályozó létrehozása többcímkes osztályozásra | 43 |
| 4.1. Irodalmi áttekintés | 45 |
| 4.2. Módszertan | 47 |
| 4.2.1. Konvolúciós neurális hálózatok | 47 |
| 4.2.2. Sigmoid veszteségfüggvény | 50 |
| 4.2.3. Hiperparaméter optimalizáció | 51 |
| 4.3. Eredmények | 53 |
| 4.3.1. Irodalmi architektúrák vizsgálta | 53 |
| 4.3.2. Röntgen specifikus súlyok létrehozása | 57 |
| 4.3.3. Hiperoptimalizálás | 58 |
| 4.3.4. Optimalizált hálózat tesztelése | 62 |
| 4.3.5. Vizualizálás | 63 |
| 4.4. Összefoglalás | 67 |
| 5. Ensemble módszerek alkalmazása többcímkes osztályozásra | 71 |
| 5.1. Módszertan | 76 |
| 5.1.1. Architektúrák | 76 |
| 5.1.2. Ensemble módszerek | 79 |
| 5.1.3. Javasolt új veszteségfüggvény | 83 |
| 5.2. Eredmények | 86 |
| 5.2.1. A különböző CNN tagok kiértékelése | 86 |
| 5.2.2. Az ensemble modellek vizsgálata | 87 |
| 5.2.3. Kombinálás transzfer tanulással | 93 |
| 5.2.4. Összehasonlítás a korszerű technikákkal | 95 |
| 5.3. Összefoglalás | 98 |
| 6. Összefoglalás | 99 |
| Köszönetnyilvánítás | 102 |
| Irodalomjegyzék | 105 |

| | |
|--------------------------------|------------|
| Függelékek | 116 |
| CITI vizsga jelentés | 116 |
| Összefoglaló | 117 |
| Publikációs jegyzék | 120 |

Ábrák jegyzéke

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1. Mellkasröntgen felvételek elváltozásokkal: (a) nincs elváltozás; (b) kardiomegália, konszolidáció, ödéma; (c) kardiomegália; (d.) atelektázia. | 4 |
| 2.1. Különböző AUC értékek és a hozzájuk tartozó ROC görbe. | 22 |
| 4.1. Folyamatábra a többlépéses optimalizációs módszerünkről. | 44 |
| 4.2. Tanulási teszt a MIMIC-CXR adatkészleten. | 56 |
| 4.3. Grad-CAM térkép az új javasolt osztályozóval. | 65 |
| 4.4. Grad-CAM térkép az alap osztályozóval. | 66 |
| 5.1. Különböző típusú együttesek: (a) osztályszintű; (b) jellemzőszintű. | 82 |
| 5.2. Módosított jellemzőszintű ensemble az új javasolt veszteségfüggvényhez. | 85 |

Táblázatok jegyzéke

| | |
|------------------------------------------------------------------------------------------------------------|----|
| 2.1. Röntgen adatbázisok áttekintése. | 8 |
| 2.2. A címkék előfordulása a ChestX-ray14 és a MIMIC-CXR adatkészletben. | 10 |
| 3.1. Általunk javasolt konvolúciós blokk bemutatása. | 36 |
| 3.2. Javasolt architektúra. | 38 |
| 3.3. Eredmények (AUC) a ChestX-ray14 adathalmazon. | 40 |
| 4.1. A modellek teljesítménye az átfogó gyorsteszten. | 55 |
| 4.2. Hiperparaméter tartományok és kiválasztott értékek. | 61 |
| 4.3. Eredmények különböző konfiguráción a ChestX-ray14 adat- készleten (átlagos AUC és szórás). | 62 |
| 4.4. Eredményeink összehasonlítása különböző tanulmányokkal. | 69 |
| 5.1. Tagok önálló teljesítménye (AUC).. | 87 |
| 5.2. Memóriahasználát és tanítható paraméterek száma a java- solt együttesek esetén. | 88 |
| 5.3. $Ensemble_1$ és $Ensemble_2$ eredmények az új javasolt vesz- teség függvény nélkül. | 90 |
| 5.4. $Ensemble_1$ eredményei az új veszteségfüggvénnyel külön- böző λ értékek mellett | 92 |
| 5.5. $Ensemble_2$ eredményei az új veszteségfüggvénnyel külön- böző λ értékek mellett | 93 |
| 5.6. Eredmények előtanított tagokkal az új és a hagyományos veszteségfüggvényt használva | 95 |
| 5.7. Eredmények összehasonlítása a legkorszerűbb eredmények- kel. | 97 |

Nyelvhasználat

A jelen disszertációban többcímű osztályozási problémával foglalkozunk mellkasröntgen felvételeken, vegyes nyelvhasználatot alkalmazva, amely magában foglalja az angol nyelv használatát is. Az alábbiakban részletezzük az alkalmazott irányelveket:

Angol szavak használata: Olyan esetekben, amelyekben nincs egyértelmű és közismert magyar megfelelője egy adott kifejezésnek vagy szónak, az angol megfelelőjét használjuk. Az ilyen angol szavak és kifejezések magyar magyarázatát aszójegyzékben adjuk meg. Például olyan szakkifejezések esetén, mint a "dense" vagy a "flatten".

Rövidítések használata: A szövegben kizárólag az angol rövidítéseket alkalmazzuk. Az összes rövidítés teljes formáját és magyar magyarázatát a rövidítések jegyzékében részletezzük. Például a "CNN" rövidítést a "convolutional neural network" helyett, és a "ROC"-t a "receiver operator characteristic" helyett.

A fentiek célja a tudományos precizitás megőrzése és az egyértelműség biztosítása, különösen azokban az esetekben, ahol a magyar terminológia nem eléggé kiforrott vagy nem teljesen egységes. Reméljük, hogy ez a megközelítés elősegíti a disszertáció tartalmának jobb megértését és egyértelműsítését.

Rövidítések jegyzéke

Adam Adaptive Moment Estimation, magyarul: adaptív momentumbecslés.

AUC Area Under Curve, magyarul: Görbe alatti terület.

BCE Binary Cross-Entropy, magyarul: Bináris Keresztentrópia.

CCE Categorical Cross-Entropy, magyarul: Kategorikus Keresztentrópia.

CNN Convolutional Neural Network, magyarul: Konvolúciós Neurális Hálózat.

FN False Negative, magyarul: Ál negatív.

FP False Positive, magyarul: Ál pozitív.

FPR False Positive Ratio, magyarul: Hamis Pozitív Arány.

Grad-CAM Gradient-weighted Class Activation Mapping, magyarul: Gradiens-súlyozott Osztály Aktivációs Térkép.

HPO Hyperparameter Optimization, magyarul: Hiperparaméter Optimalizáció.

LSTM Long Short-term Memory, magyarul: hosszú rövidtávú memória.

RMSProp Root Mean Square Propagation, magyarul: Négyzetes Közép Terjedés.

ROC Receiver Operating Characteristic, magyarul: Vevő működési karakterisztika.

SFL Sigmoid Focal Loss, magyarul: Szigmoid Fókuszú Veszteség.

SGD Stochastic Gradient Descent, magyarul: stokasztikus gradiens leereszkedés.

TN True Negative, magyarul: Valós negatív.

TP True Positive, magyarul: Valós pozitív.

Szójegyzék

attention learning A figyelem tanulás egy olyan technika a gépi tanulásban és mélytanulásban, amely lehetővé teszi a modell számára, hogy a bemeneti adatok különböző részeire eltérő mértékben összpontosítson, javítva ezzel a teljesítményt és az értelmezhetőséget.

bias A neurális hálózatokban egy olyan paraméter, amelyet a bemenetekhez adunk hozzá minden egyes neuronnál a lineáris kombináció részeként, javítva a modell tanulási képességét .

dense A sűrű réteg, más néven teljesen összekapcsolt réteg (dense layer), egy neurális hálózat rétege, ahol minden neuron kapcsolódik az előző réteg összes neuronjához, és amely a bemeneti jellemzők és a kimeneti osztályok közötti nemlineáris összefüggéseket tanulja meg.

dropout A dropout egy regularizációs technika, amelyet a neurális hálózatok tanulása során alkalmaznak a túlillesztés elkerülésére. A dropout során véletlenszerűen kikapcsolják a hálózat egyes neuronjait az edzés folyamán, ezáltal csökkentve a neurális hálózatok közötti kölcsönhatásokat. Ez a technika javítja a hálózat általánosítási képességét, mivel megakadályozza, hogy a hálózat túlságosan függjön egy adott neurontól vagy jellemzőtől. A dropout mértékét egy valószínűségi érték p határozza meg, amely azt jelzi, hogy egy neuron milyen valószínűséggel lesz kikapcsolva.

ensemble Az együttes modell egy olyan gépi tanulási megközelítés, amely több különálló modellt kombinál annak érdekében, hogy javítsa az osztályozás pontosságát és megbízhatóságát. Az együttes modellek jellemző szintű és osztály szintű kombinációkat is alkalmazhatnak.

flatten A flatten réteg egy neurális hálózat komponense, amely átalakítja a többdimenziós bemenetet egy egydimenziós vektorrá, ezáltal lehetővé téve a konvolúciós és teljesen összekapcsolt rétegek közötti adatátvitelt.

kernel A kernel, vagy szűrő, a konvolúciós neurális hálózatokban (CNN-ekben) használt mátrix, amely a bemeneti adatokon végighaladva, azaz konvolválva, jellemzőket nyer ki. A kernel egy kisméretű, általában 3×3 vagy 5×5 dimenziójú mátrix, amelyet a konvolúciós réteg súlyai alkotnak. A kernel

célja, hogy az input képen mintázatokat, például éleket, textúrákat és más fontos jellemzőket detektáljon.

pooling A pooling réteg egy neurális hálózat komponense, amely a bemeneti jellemzőtérképek méretét csökkenti és a legfontosabb jellemzőket emeli ki, miközben csökkenti a hálózat számítási terhelését és segít a túlillesztés elkerülésében. Két leggyakoribb típusa a max pooling és az átlag pooling

softmax A softmax egy aktivációs függvény, amely az egyes kimeneti neuronjainak értékeit valószínűséggé alakítja, amelyek összege 1, lehetővé téve az osztályok közötti valószínűségek kiszámítását.

state-of-the-art A state-of-the-art kifejezés arra utal, hogy egy technológia vagy módszer a jelenlegi legfejlettebb és legkorszerűbb a maga területén, a legjobb teljesítményt nyújtva és a legújabb kutatási eredményeken alapulva.

1. fejezet

Bevezetés

A mellkasröntgen felvételek alapvető szerepe az egészségügyben vitathatatlan, mivel számos diagnosztikus folyamat alapjaként szolgálnak. Széleskörű használatukat számos előny támogatja: alacsony sugárterhelés, a kis kórházaktól a nagy központi klinikákig terjedő elérhetősége, gyors és olcsó előállíthatóság[1]. Ezek a tulajdonságok teszik a mellkasi röntgenfelvételeket a különböző alapbetegségek, például tüdőfertőzések, tüdőgyulladások, szívproblémák és bizonyos ráktípusok diagnosztizálásának alapvető eszközévé. Az Egészségügyi Világszervezet szerint évente több mint négy-millió ember meghal tüdőgyulladásban, ami az évi 450 millió tüdőgyulladós eset közel 0,88%-os halálozási arányát jelenti. Bár a halálozási arány alacsonynak tűnhet, riasztó, hogy a tüdőgyulladás világszinten az összes halálozás 7%-áért felelős. Különleges jelenség is kíséri a tüdőgyulladást,

mivel a fejlődő országokban a gyermekeknél, ezzel szemben a fejlett országokban pedig az idősebb betegeknel jelent halálos fenyegetést [2].

A mellkasröntgen felvétel az egyik leggyakoribb orvosi képalkotó vizsgálat a világon. Fontosságára való tekintettel várható, hogy a mellkasi röntgenfelvételek a közeljövőben is a leggyakrabban igényelt diagnosztikai vizsgálatok között lesznek világszerte. Nem lehet eléggé hangsúlyozni, hogy a gyors információnyújtás hatékonysága miatt nélkülözhetetlen a sürgősségi osztályokon, intenzív osztályokon és a rendszeres egészségügyi szűrővizsgálatokon.

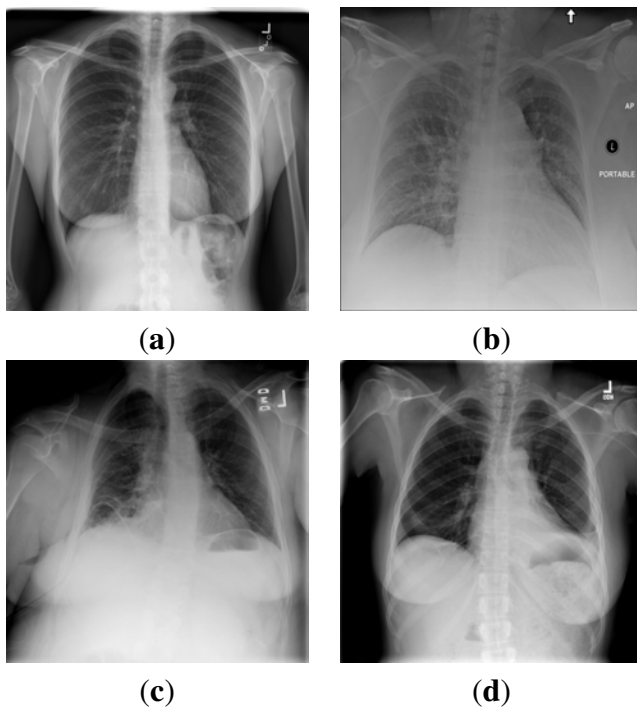
A mellkasröntgen felvételek értelmezése azonban számos kihívással jár. Olyan szintű tudást és szakértelmet igényel, amellyel jellemzően csak a képzett radiológusok rendelkeznek. Ez az igény akadályt jelent az egészségügyi rendszerekben, különösen a képzett szakemberek hiányával küzdő területeken. Amíg az Egyesült Államokban évről évre nő a radiológusok száma [1]; addig vannak olyan országok, például Libéria, ahol súlyos a radiológushiány: 2015-ben négy millió emberre mindössze két aktív radiológus szakember jutott [3]. Mivel a képek alapos elemzése elengedhetetlen az orvosi értelmezéshez, ez erősen korlátozhatja a skálázhatóságot, hiába válik maga a vizsgálat évről évre olcsóbbá és gyorsabbá. A szakember hiány késedelmet okozhat a betegek diagnosztizálásában és kezelésében.

A sürgősségi osztályok a mellkas-röntgenvizsgálatok kiemelt helyszínei, a jelentések szerint a sürgősségi osztály látogatások 34,4%-ában hasz-

nálnak valamilyen mellkas-röntgenvizsgálatot [4]. Így például a COVID-19 világjárvány idején jelentősen megnőtt a radiológia alkalmazásának gyakorisága a tüdőgyulladás diagnózisában. Viszont a radiológusok száma nem skálázható ilyen gyorsan. Az általános orvosi diploma megszerzése után még legalább 5 év, hogy valakiből radiológus szakorvos lehessen. A humán erőforrás-korlátok válsághelyzetben nem oldhatók meg gyorsan, a szakmákhoz szükséges többéves képzés miatt. A képalkotó vizsgálatok növekvő használata a sürgősségi ellátásban kritikus problémává vált a humán erőforrás korlátozottsága miatt.

A mellkasröntgen felvételek számítógéppel történő elemzése összetett kihívást jelent, mivel gyakran többféle patológiát tartalmaznak. Ami egy különleges többcímkes osztályozási problémához vezet. Például a jelen disszertációban is használt ChestX-ray14 adatkészlet[5], minden egyes képet több, főként a tüdőt és a szívet érintő patológiával annotáltak (ld. 1.1. ábra), számszerűen 14 különböző elváltozást címkéz. A korábbi tanulmányokban (amelyeket később részletezünk) a különböző patológiák egyenlő súlyt kapnak az osztályozási folyamatban, és minden betegséget azonos fontossággal kezelnek a röntgenképek címkézésekor.

Ezzel szemben meg kell jegyeznünk, hogy a címkék között gyakran léteznek összefüggések, ilyen például a kardiomegália jelenléte, amely gyakran a tüdőödéma előfordulásának nagyobb kockázatával jár együtt, ami arra utal, hogy a patológiák közötti összefüggések feltárása és megértése kulcs-



1.1. ábra. Mellkasröntgen felvételek elváltozásokkal: **(a)** nincs elváltozás; **(b)** kardiomegália, konszolidáció, ödéma; **(c)** kardiomegália; **(d.)** atelektázia.

fontosságú lehet az osztályozási pontosság javításához. Míg azonban ezek a korrelációk bizonyos esetekben segíthetnek a diagnózis tisztázásában, a nem kapcsolódó címkék zajt vihetnek a rendszerbe, ami megnehezíti az osztályozó számára a releváns jellemzők azonosítását és a hatékony tanulást. Ebben a kutatásban kizárólag a képadatokra összpontosítunk, és nem használunk semmilyen további külső információt, például a beteg nemét, korát vagy a fókuszálást segítő kiemelt részek koordinátáit, amelyekről más

tanulmányokban kimutatták, hogy javítják a osztályozó modell teljesítményét.

A számítógéppel támogatott diagnosztika, ezen belül is az orvosi képfeldolgozás: a képi jellemzők kinyerésén alapul. Egy hatékony algoritmus megalkotása a hagyományos technikák alkalmazásával szoros együttműködést igényel a radiológusokkal, akik egy adott elváltozás különböző paramétereit, például az alakot, a méretet vagy a színt (azaz a sűrűséget) megadják a képfeldolgozó szakembereknek, hogy ezek detektálása utána egy kritériumalapú döntési fát építhessenek fel.

Ez a technika, igaz, csak néhány darab mintát igényel, de a radiológus szakemberek részéről hatalmas munkaóra mennyiséggel jár [6, 7, 8]. Ezen fát gyakran gépi tanulási algoritmussal helyettesítik vagy egészítik ki. Ez az eljárás a fejlesztési folyamat során kevesebb órányi ráfordítást igényel az orvostól. A folyamathoz viszont sok, helyesen annotált képre van szükség ahhoz, hogy az eddig az orvosok által megadott paraméter határokat automatikusan, felügyelt tanulási technikák segítségével meghatározzuk [9].

Ezen disszertáció a jelenleg legfejlettebb megoldást használja: a mélytanulási hálózatokat. Az ilyen hálózatok alkalmazásakor a képi jellemzők kinyerése is a tanulási folyamat része, nem igényel közvetlen szakorvosi együttműködést, csak több tízezer annotált képet.

2. fejezet

Alapkonceptiók és módszerek

2.1. Használt adatok

Számos adatbázist böngésztünk át a kiválasztási folyamat során (ld. 2.1. táblázat). Kutatásunk alapját végül két jelentős, hozzáférhető mellkasi röntgenfelvétel adatbázis képezte, amelyek kiválasztását szigorú kritériumok alapján végeztük el, figyelembe véve azok hozzáférhetőségét, a képek számát és a címkék többszörös jellegét. Ezek a szempontok a National Institutes of Health által kiadott ChestX-ray14 adatbázishoz és a Massachusetts Institute of Technology által összeállított MIMIC-CXR adathalmazhoz vezettek minket, amelyek együttesen kiváló alapot nyújtottak a mélytanulási algoritmusok orvosi képeken való alkalmazásának kutatásához.

2.1. táblázat. Röntgen adatbázisok áttekintése.

| Adatkészlet neve | Betegek száma | Képek száma | Címkék száma | Képek formátuma | Képek mérete |
|---------------------|---------------|-------------|--------------|-----------------|----------------------------|
| ChestX-ray14 [5] | 30 805 | 112 120 | 14 | PNG | 1024 × 1024 |
| MIMIC-CXR [10, 11] | 65 383 | 371 920 | 13 | DICOM, JPEG | 3520 × 4280 1377 × 1153 |
| CheXpert [12] | 65 240 | 224 316 | 14 | DICOM | nincs adat |
| PLCO [13] | 56 071 | 185 421 | 12 | DICOM | n.a. |
| PadChest [14] | 67 525 | 160 868 | 193 | n.a. | n.a. |
| Ped-pneumonia [15] | n.a. | 5856 | 2 | n.a. | n.a. |
| RSNA-Pneumonia [16] | n.a. | 30 000 | 1 | DICOM | n.a. |
| BIMCV [17] | 9129 | 25 554 | 1 | DICOM | n.a. |

2.1.1. ChestX-ray14

A ChestX-ray14 adatbázis különösen figyelemre méltó, 112 120 frontális nézetű röntgenfelvétellel, amelyeket 14 különböző típusú mellkasi patológiával címkéztek, beleértve a tüdőgyulladást, az ödémát és a pneumothoraxot. Ez az adathalmaz lehetővé teszi a kutatók számára, hogy széles körben teszteljék és validálják a képfeldolgozó algoritmusokat valóságos orvosi képek sokféleségén.

2.1.2. MIMIC-CXR

A MIMIC-CXR, amely még nagyobb adathalmaz, 371 920 képet tartalmaz, köztük frontális és laterális felvételeket, így mélyebb betekintést nyújt a mellkast érintő betegségek diagnosztikába és szélesebb spektrumú patológiákba. Az adatbázist a PhysioNet[18] keresztül értük el, amelyhez vizsga szükséges, ezen követelményt, a CITI Program - Human Research - Data or Specimens Only Research vizsgát 100%-os eredménnyel teljesítet-

tük 2019. február 14-én (ld. melléklet - CITI vizsga jelentés).

2.1.3. Adatok összehasonlítása

Mindkét adathalmaz a mélytanulás területén, különösen az orvosi képelemzési kutatások számára nélkülözhetetlenné vált. A ChestX-ray14 adathalmaz a röntgen alapú mélytanulási modellek összehasonlításának egyik alappillére lett. A MIMIC-CXR adathalmaz, amelyet a Beth Israel Deaconess Medical Centerrel együttműködve gyűjtöttek, további adatokkal szolgál a kutatásokhoz, egy hatalmas és változatos adatforrást kínál a modellek képzéséhez és teszteléséhez.

Eredetileg a MIMIC-CXR adatbázison szerettünk volna kutatásokat végezni, de nem áll rendelkezésre elég publikált eredmény ahhoz, hogy összehasonlíthatóak legyenek a eredmények, ezért a MIMIC-CXR adatbázis jelenleg kizárólag a súlyok előtanítására használjuk. A ChestX-ray14 és a MIMIC-CXR adathalmazok használatával végzett tanulmányaink hozzájárulnak a mellkasröntgenképek automatizált elemzésének fejlődéséhez, lehetővé téve az orvosok számára, hogy pontosabb diagnózist állíthassanak fel és hatékonyabban kezelhessék a betegeket. Az adatbázisokban szereplő címkék előfordulását a 2.2. táblázat tárgyalja.

Ahogy a a 2.2. táblázat mutatja az egyes címkék előfordulása között számottevő különbség van, ezeket az eltérő előfordulásait különféle módszerekkel lehetne kezelni, például címke-specifikus súlyozással vagy ki-

2.2. táblázat. A címkék előfordulása a ChestX-ray14 és a MIMIC-CXR adatkészletben.

| Címke neve | Betegség neve | ChestX-ray14 | MIMIC-CRX |
|--------------------------|----------------------------|---------------------|------------------|
| Airspace opacity | légtartalom csökkenés | – | 55 660 |
| Atelectasis | atelektázia | 8280 | 49 638 |
| Cardiomegaly | kardiomegália | 1707 | 48 899 |
| Consolidation | konzolidáció | 2852 | 11 734 |
| Edema | tüdődéma | 1378 | 29 390 |
| Effusion | effúzió | 8659 | – |
| Emphysema | emfizéma | 1423 | – |
| Enlarged cardiomeastinum | mediastinalis kiszélesedés | – | 7868 |
| Fibrosis | fibrózis | 1251 | – |
| Fracture | törés | – | 5018 |
| Hernia | rekeszsérv | 141 | – |
| Infiltration | tüdőinfiltrátum | 13 782 | – |
| Lung Lesion | tüdőgóc | – | 7003 |
| Mass | térfoglalás / terime | 4034 | – |
| Nodule | göb | 4708 | – |
| Pleural effusion | mellúri folyadékgyülem | – | 58 734 |
| Pleural other | egyéb mellhártya eltérés | – | 2137 |
| Pleural thickening | mellhártya megvastagodása | 2242 | – |
| Pneumonia | tüdőgyulladás | 876 | 18 330 |
| Pneumothorax | légmell | 2637 | 11 610 |
| Support Devices | orvosi eszközök | – | 74 247 |

egyensúlyozott mintavétellel. Ezek a technikák lehetővé tennék a ritkábban előforduló címkék jobb figyelembevételét, javítva az osztályozási teljesítményt. Azonban jelen dolgozatban nem alkalmaztunk ilyen módszereket, mivel az alapértelmezett, címkefüggetlen megoldásokra fókuszáltunk, hogy a módszerek általános alkalmazhatóságát és hatékonyságát vizsgáljuk.

2.1.4. Címkék reprezentálása

Mint korábban említettük, a jelenlegi kontextus a röntgenképek többcímkes osztályozása. Ez a bináris osztályozás egy speciális területe. A bináris osztályozás esetén egy modell két lehetséges kimenetet különböztet meg, ami jellemzően egy adott osztály jelenlétét vagy hiányát jelenti. A többcímkes osztályozás ezt a logikát terjeszti ki, mivel minden egyes osztályra külön bináris osztályozót alkalmaz, lehetővé téve, hogy egy minta egyszerre több osztályhoz is tartozzon. Így a többcímkes osztályozás alapja a bináris osztályozás, de bonyolultabb, mivel több osztályozási feladatot kombinál egy modellben. Ha K címkével rendelkezünk, ez azt jelenti, hogy minden képnek van K darab független címkéje. A többcímkes osztályozásban egy célvektor egy K -dimenziós $T = [t_1, t_2, \dots, t_K]$ vektorral reprezentálunk, ahol K a címkék számát, t_n az adott címkét jelenti.

$$t_k = \begin{cases} 1, & \text{ha az } k\text{-adik címke jelen van a képen} \\ 0, & \text{ha az } k\text{-adik címke nincs jelen a képen} \end{cases} . \quad (2.1)$$

2.1.5. Adatok orvosi háttere

A **tüdőinfiltrátum** felhalmozódott kóros anyag jelenléte a tüdőben. Mivel nem specifikus kifejezésnek számít, ezért további differenciálás javasolt pl. tüdőgyulladás, az atelektázia, a vérzés, az ödéma, a fibrózis irányába.

Az **atelektázia** egy kóros állapot, amikor a tüdő egy részének légtartalma felszívódik, az adott egyes terület nem tágul ki megfelelő mértékben. A mellkasröntgen felvételeken az atelektázia lineáris árnyékokkal (fehérséggel), zsúfolt léghörgőgrammokkal és a tüdőerek zsúfoltságával jelentkezik. Jellemzően csökken a tüdő térfogata is.

A **pneumothoraxban** vagyis légmellben, szemben az atelektáziával, nemcsak a tüdő egyes részei, hanem az egész tüdő részben vagy teljesen összeesik, a pleurális térbe jutó levegő miatt. A sima röntgenfelvételen az alapján azonosítható, hogy vékony, éles fehér vonalként láthatóvá válik a visceralis pleurális él és ettől a vonaltól a tüdő szélé irányába már nem látható a tüdő normális rajzolata,

A **konzolidáció** jelenségét látjuk, amikor az alveoláris légterek folyadékkal, gennyel vagy sejtekkel töltődnek. A röntgenfelvételeken a röntgensugarak fokozott elnyelődése miatt az adott területen megnövekedett fehérséget láthatunk. Fontos a konzolidáció pontos helyének meghatározása a tüdőgyulladás, az ödéma, a neoplasia és az embólia meghatározásában.

A **tüdőödéma** a folyadék kóros felhalmozódása a tüdőben. A mellkasröntgen-felvétele több jel kombinációit mutatja, mint például a pleurális folyadékgyülem, az egyes területek homályosodása a denevérszárny mintázatban eloszlásában és a kardiomegália.

Az **effúzió** a többletfolyadék felhalmozódása egyes terekben, például a tüdő körüli pleurális térben (pleurális folyadékgyülem) vagy a szív körüli

szívburk térben (perikardiális folyadékgyülem). A rutin mellkasröntgenen a costophrenicus vagy cardiophrenicus szög eltompulását, esetleg meniszkuszképződést észlelünk.

A **kardiomegália** a szív megnagyobbodását jelenti, amelynek számos oka lehet. A kórosan megnagyobbodott szív sziluettjét könnyen észlelhetjük mellkasröntgen felvételeken, kiszámíthatjuk a szív-mellkas arányát, mely a szív sziluettjének szélessége osztva a mellüreg szélességével, amely normál esetben kisebb, mint 0,5.

A **tüdőgyulladás** egy általános kifejezés a tüdő gyulladásos állapotára, amely elsősorban az alveolusokat érinti. A röntgenfelvételek a tüdőgyulladás különböző típusától függően változnak. Vírusos tüdőgyulladásban normális vagy foltos megjelenést mutathat, mely egy vagy mind a két oldalon megjelhet. Egyes bakteriális eredetű tüdőgyulladás esetén akár egy teljes lebeny konszolidációját is láthatjuk. Aspirációs pneumoniában leggyakrabban a jobb alsó lebeny érintett.

Az **emfizéma** az alveolusok kóros megnagyobbodása, amelyet az alveoláris fal pusztulása kísér. Előrehaladott állapotban már mellkasröntgenfelvételen is láthatóvá válnak a nagy léghólyagok, ún. bullák képződése. Ugyanakkor enyhébb esetekben közvetett jelekre támaszkodunk, pl. a röntgenkép általános fényesség növekedésére, amelyet a széles bordák és a laposan, alsó helyzetben lévő rekeszizom kísér, melyek mind a tüdő túltágulásának jelei.

A **mellhártya megvastagodása** jóindulatú (például visszatérő gyulladás, pleurális empyema) vagy rosszindulatú állapotok (például metasztázis, mesothelioma, limfóma) következménye is lehet. A mellhártya megvastagodása lapos, szabályos, szabálytalan vagy ék alakú opacitást képez a röntgenfelvételeken, 5 mm-nél kisebb szélességben, és leggyakrabban a tüdő csúcán jelenik meg.

A **fibrózis** a tüdőben lévő fibrotikus szövet felszaporodása, amely a perifériás légutak patológiás tágulásával és a tüdő térfogatának csökkenésével járó parenchimális destrukcióval jár együtt. A röntgenfelvételen a tüdő perifériáin (szélei) közelében fibrotikus-hálózatos rajzolatot láthatunk, árnyékolás és a szív határainak bolyhosodása is megfigyelhető.

A **rekeszsérv** a gyomor egy részének a rekeszizmon keresztül a mellkasba történő kidudorodása a nyelőcsőnek fenntartott résen keresztül. Nagyméretű sérv esetén a gyomorban lévő gáz/folyadék szintje is látható a mellkasröntgenen, ami egy könnyen felismerhető egyenes vízszintes vonalat hoz létre, és gyakran a szív árnyékát fedi.

A tüdőben lévő **göbök** (nodulusok) és térfoglaló folyamatok rendellenes növekedések. Ez lehet jól vagy rosszul körülhatárolt, diszkrét, kerekded homályosság is. A göcök 3 cm vagy annál kisebb átmérőjű térfoglalások, míg a daganatok a 3 cm-nél nagyobbak. A nodulusok számát, méretét, morfológiáját és eloszlását is fontos értékelni, mivel ezek segíthetnek a differenciáldiagnózis felállításában.

A mellkasröntgen felvételeken láthatóak lehetnek még friss és már gyógyulófélben lévő bordatörés(ek), kulcscsonttörés jelei is.

2.2. Definíciók

Ebben a fejezetben definiáljuk a dolgozat elkészítéséhez szükséges fogalmakat és eljárásokat. A fejezet célja, hogy az olvasó megértse és kövesse a dolgozatban használt módszereket és technikákat.

2.2.1. Digitális kép

1. Definíció. Legyen I egy szürkeskálás digitális kép, amely egy $W \times H$ méretű rácson értelmezett függvény. A függvény $I : D \rightarrow V$ minden képpontot (pixel) egy szürkeárnyalat-értékkel rendel össze, ahol:

$D = \{(x, y) \mid x = 0, 1, 2, \dots, W - 1; y = 0, 1, 2, \dots, H - 1\}$ a képpontok halmaza egy W szélességű és H magasságú képen, $V = \{0, 1, 2, \dots, 255\}$ a lehetséges szürkeárnyalat-értékek halmaza a standard 8 bites szürkeskálás képek esetén.

Minden $(x, y) \in D$ képponthoz $I(x, y) \in V$ rendel egy értéket, ami a pixel szürke árnyalatát jelöli, kifejezve ezzel a pixel luminanciáját vagy fényerejét. Az $I(x, y)$ értékek leképezése lehetővé teszi a kép vizuális tartalmának reprezentálását színek nélkül, kiemelve a fény- és árnyékhatásokat, amelyek fontos információkat szolgáltatnak a képfeldolgozási feladatokban.

2.2.2. Képaugmentáció

A képaugmentáció egy adatbővítési technika a gépi tanulás és különösen a mélytanulás területén, amely a meglévő képadatok módosításával növeli az adathalmaz méretét és diverzitását anélkül, hogy új, valódi képeket kellene gyűjteni. Ez a technika javítja a modellek általánosítási képességét azáltal, hogy azokat többféle képanyagon is tanítja, így csökkenti a túlilleszkedés esélyét. A képaugmentáció matematikailag az alábbi módon írható le:

2. Definíció. Legyen \mathcal{I} az eredeti képek halmaza. Egy augmentációs függvény, A , az $I \in \mathcal{I}$ képekhez új képeket generál az alábbi transzformációk alkalmazásával:

$$A : \mathcal{I} \rightarrow \mathcal{I}$$

Az A függvény által alkalmazott transzformációk közé tartozhatnak:

- Geometriai transzformációk: például forgatás, méretezés (skálázás), eltolás (transzláció), tükrözés (horizontális vagy vertikális).
- Színtér-transzformációk: például színinverzió, fényerő módosítása, kontraszt állítása.
- Zajadagolás: különböző típusú zaj (pl. Gauss-zaj) hozzáadása a képekhez. Képmanipuláció: élesítés, homályosítás, vágás.

Az A által előállított képek, $A(I)$, bemenetként szolgálnak a tanulási modellhez, így bővítve az eredeti adathalmazt új, változatos képekkel, amelyek a valós világ változatos viszonyait modellezik. Ezzel a módszerrel a tanulási modell nemcsak az eredeti, hanem az augmentált képek jellemzőire is optimalizálódik, ezáltal növelve a modell robusztusságát és általánosítási képességét. Természetesen ezen függvények sorozata is lehet egy kép augmentáció, így egyszerre lehet többféle transzformációt is használni a képeken.

2.2.3. Bináris osztályozás

A digitális képek bináris osztályozása azt a folyamatot jelenti, amely során minden képet két előre meghatározott osztály (például osztály 0 vagy osztály 1) egyikébe sorolunk a kép vizuális tartalmát reprezentáló jellemzők alapján. Matematikailag ezt a folyamatot a következőképpen írhatjuk le:

3. Definíció. Legyen \mathcal{I} az összes lehetséges digitális kép halmaza, és $\mathcal{C} = \{0, 1\}$ jelölje a két osztályt. Egy bináris osztályozó $f : \mathcal{I} \rightarrow [0, 1]$ ami egy összetett függvény $f = g \circ \phi$, ahol:

$\phi : \mathcal{I} \rightarrow \mathbb{R}^n$ egy jellemzőleképezés, amely minden képet I átalakít egy n -dimenziós jellemzővektorra $\phi(I)$.

$g : \mathbb{R}^n \rightarrow [0, 1]$ egy osztályozó függvény, amely a képi jellemzők $\phi(I)$

alapján megad egy valószínűséget annak függvényében, hogy a képi jellemzők az 1-es osztályba tartoznak-e.

2.2.4. Többcímkes osztályozás

A digitális képek többcímkes osztályozása egy olyan folyamat, amely során minden képet több, egymástól függetlenül értelmezett osztály egyikébe vagy többébe sorolunk a kép vizuális tartalmát reprezentáló jellemzők alapján. Matematikailag ezt a folyamatot a következőképpen írhatjuk le:

4. Definíció. Legyen \mathcal{I} az összes lehetséges digitális kép halmaza, $\mathcal{C} = \{1, 2, \dots, K\}$ jelölje a K lehetséges osztályt, és minden kép $I \in \mathcal{I}$ esetén legyen $\mathbf{y}_I \in \{0, 1\}^K$ a címkek bináris vektora, ahol minden y_{Ik} azt jelzi, hogy a kép I tartozik-e az k -adik címkéhez ($y_{Ik} = 1$) vagy sem ($y_{Ik} = 0$). A többcímkes osztályozó egy függvény f :

$$f : \mathcal{I} \rightarrow [0, 1]^K$$

ahol f minden képhez rendel egy K -dimenziós valószínűségi vektort, $\mathbf{p}_I = (p_{I1}, p_{I2}, \dots, p_{IK})$. Minden p_{Ik} a valószínűség, hogy az I kép tartozik az k -adik osztályhoz. Az osztályozás megvalósítása általában a következő lépéseket foglalja magában:

$\phi : \mathcal{I} \rightarrow \mathbb{R}^n$ egy jellemzőleképezés, amely minden képet átalakít egy n -dimenziós jellemzővektorrá.

$g : \mathbb{R}^n \rightarrow [0, 1]^K$ egy döntési függvény, amely a jellemző vektor alapján megadja minden osztályra a hozzá tartozó valószínűséget.

2.2.5. Metrikák

A ROC görbe és az AUC kritikus eszközök a bináris osztályozó modellek teljesítményének értékelésére. Gyakran hatékonyabb és informatívabb választás, mint az egyszerű pontosság, különösen olyan helyzetekben, ahol a modell tanulási képességének az összehasonlítása a fontos, mivel nem függ az osztályozási küszöbtől. Mivel a többcímkes osztályozás esetén az egyes címkék úgy viselkednek mint ha egy-egy bináris osztályozó kimenetei lennének, ezért címkénként alkalmazhatjuk rájuk a ROC és az abból számolt AUC metrikát. Ezek a metrikák különösen fontosak az orvosi diagnosztika és a gépi tanulás területein, ahol az érzékenység és a specifitás egyensúlya döntő jelentőségű. Érzékenység (Szenzitivitás): Az érzékenység a pozitív esetek helyes azonosításának arányát jelzi.

$$\text{Szenzitivitás} = \frac{TP}{TP + FN} \quad (2.2)$$

ahol TP a valódi pozitívok száma, FN pedig a hamis negatívok száma. Ez azt mutatja meg, hogy a modell milyen jól ismeri fel a valódi pozitív eseteket. Specifitás: A specifitás a negatív esetek helyes azonosításának arányát

jelzi.

$$\text{Specifitás} = \frac{TN}{TN + FP} \quad (2.3)$$

ahol TN a valódi negatívok száma, FP pedig a hamis pozitívok száma. Ez azt mutatja meg, hogy a modell milyen jól ismeri fel a valódi negatív eseteket. A ROC görbe egy grafikon, amely az érzékenység az hamis pozitív arány (FPR) függvényében amely jelen levezetésben $1 - \text{Specifitás}$ ábrázolja az összes lehetséges döntési küszöbérték mellett.

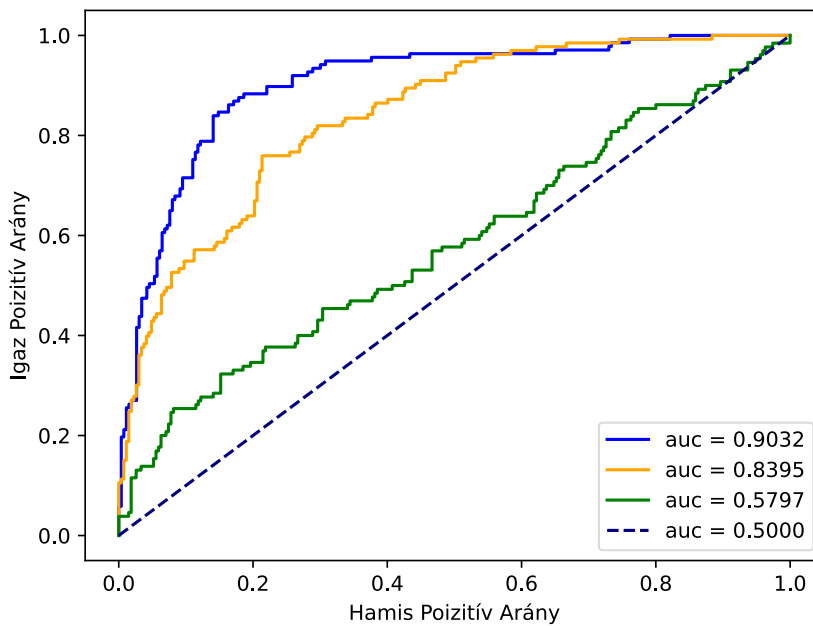
Az AUC értéke, azaz a ROC görbe alatti terület, kvantitatív mérőszámot ad arra, hogy az adott osztályozó milyen jól képes megkülönböztetni a két osztályt. Az AUC értéke a teljes tökéletességtől (1,0) a teljes véletlenszerűségig (0,5) terjedhet. Az ideális osztályozó AUC értéke 1, ami azt jelenti, hogy tökéletesen képes megkülönböztetni a két osztályt minden lehetséges küszöbérték mellett.

A ROC görbe ábrázolása segít vizualizálni a döntési küszöbök változtatásának hatását a modell érzékenységére és 1-specifitására. Az AUC értéke minél közelebb van 1-hez, annál jobb az osztályozó teljesítménye. Ezzel szemben, egy AUC érték, ami 0,5-höz közeli, azt jelzi, hogy az osztályozó teljesítménye nem jobb a véletlen tippelésnél (ld. 2.1. ábra).

Jelen dolgozatban az AUC metrikát, úgy alkalmaztuk többcímkes környezetben, hogy minden egyes címkét külön-külön kiértékelünk, majd az egyes címkék AUC értékeit átlagoljuk. Ez a módszer biztosítja, hogy minden címke külön figyelembe vételre kerüljön az osztályozási teljesítmény

értékelésekor. Az így kapott átlagos AUC érték jobb képet ad az egész modell általános teljesítményéről a különböző címkék között, mivel az esetleges egyéni címkék eltéréseit is kiegyenlíti.

Jelen dolgozatban az AUC értéket használjuk a saját megoldásaink és a legkorszerűbb megoldásokkal való összehasonlítások során is.



2.1. ábra. Különböző AUC értékek és a hozzájuk tartozó ROC görbe.

2.2.6. Aktivációs és veszteségfüggvények

A legtöbb képalapú osztályozási probléma bináris vagy kategorikus, nem pedig többcímkes osztályozás. A bináris osztályozás esetén két lehetséges kimeneti kategória van, míg a kategorikus osztályozásnál több lehetséges, egymást kizáró kategória létezik, de egy minta csak egy kategóriához tartozhat. Ezzel szemben a többcímkes osztályozás lehetővé teszi, hogy egy minta egyszerre több kategóriához is tartozzon, így minden kategóriához egy külön kimeneti címkét rendel. A főbb keretrendszerek nem rendelkeznek egyértelmű dokumentációval a többcímkes problémákra. Az első két típus egyértelmű: A bináris osztályozás az elemek két különböző csoportba sorolásának módszere (ld. 3. definíció). Ezeknek a modelleknek kimenetén a végső aktivációs függvény egy szigmoidot (2.4), veszteségfüggvényként pedig bináris keresztentropiát (BCE) kell használnunk.

$$f(s_i) = \frac{1}{1 + e^{-s_i}} \quad (2.4)$$

A BCE egy széles körben használt veszteségfüggvény bináris osztályozási feladatokhoz. A BCE a valódi címkék és a modell által jósolt valószínűségi értékek közötti különbséget méri.

A BCE veszteség egy adatpontra a következőképpen adódik:

$$L_{BCE} = -y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}), \quad (2.5)$$

ahol y a valódi címke (0 vagy 1), \hat{y} pedig annak a valószínűsége, hogy az osztálycímke 1 lesz.

Az említett másik gyakran használt osztályozási mód a kategorikus osztályozási probléma, itt az elemek K nem átfedő csoportba sorolása történik. A kategorikus modellek softmax 2.6 aktivációs függvényt használnak és kategorikus keresztentropiát (CCE) veszteségfüggvényt.

$$f(s_i) = \frac{e^{s_i}}{\sum_j^K e^{s_j}} \quad (2.6)$$

A CCE egy általánosan használt veszteségfüggvény többosztályos osztályozási feladatokhoz. A BCE-hez hasonlóan arra szolgál, hogy számszerűsítse a két valószínűségi eloszlás közötti eltérést; azaz a tényleges osztálycímkeket reprezentáló valós eloszlást és a modell által generált előre jelzett eloszlást. A CCE veszteség egyetlen adatpontra a következőképpen fejezhető ki:

$$L_{CCE} = - \sum_{c=1}^K y_c \log(\hat{y}_c) \quad (2.7)$$

Itt C az osztályok számát jelöli, y_c az egyes osztályok valódi valószínűségi értékét, \hat{y}_c pedig az adott osztályba tartozó adatpont előre jelzett valószínűségét. A CCE értékes eszköz, a több mint két különböző osztályt tartalmazó feladatok modelljeinek képzésében.

Ezért mivel a többcímkes osztályozás, több független bináris címke

együttesét jelenti (ld. 4. definíció). Ezért az általunk továbbiakban használt veszteségfüggvényt a BCE függvényből vezetjük le. Ebben a módosításban a veszteségfüggvényt minden címkére függetlenül alkalmaztuk. Ez a független kezelés lehetővé teszi, hogy a modell minden egyes címkét külön bináris osztályozási problémaként tanuljon. A többcímkes osztályozáshoz adaptált BCE veszteséget egy adatpontra a következőképpen számítjuk ki:

$$L_{Multi-BCE} = -\frac{1}{K} \sum_{i=1}^K [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2.8)$$

Itt K a független címkék száma, y_i az i -edik címke valódi címkéje (0 vagy 1), és \hat{y}_i az adott címke 1-es valószínűségének előrejelzett valószínűsége. A veszteséget az összes címkére átlagoljuk, s figyelembe vesszük a címkék példányonkénti változó számát. Ez az adaptáció lehetővé teszi a BCE használatát többcímkes forgatókönyvekben, így sokoldalú eszközzé válik az osztályozási feladatok széles körének kezelésére. Az (2.8) egyenletből láthatjuk, hogy ez az osztályozási probléma minden egyes címkére független bináris osztályozásként ábrázolható. Természetesen aktivációs függvényként ugyancsak a szigmoidot alkalmazzuk.

2.3. Algoritmusok

Ebben a fejezetben bemutatjuk a dolgozatomban, használt algoritmusokat. Ez magában foglalja a mélytanulás során használt optimalizálókat és az általunk használt keresztvalidációs technikát.

2.3.1. SGD

A SGD egy alapvető optimalizáló algoritmus (1. algoritmus), amelyet arra használnak, hogy iteratív módon finomítsa a modell paramétereit az adathalmazon. Bár lehetővé teszi a gyorsabb konvergenciát nagy adathalmazok esetén, érzékeny a lokális minimumokra, ami megakadályozhatja a globális optimum elérését. Ezzel szemben a momentumos SGD egy továbbfejlesztett változat, amely a hagyományos sztochasztikus csökkenő gradiens módszer előnyeit kombinálja a mozgó átlagolás módszerével. Ez csökkenti a lokális minimumokba való beragadás kockázatát, és gyakran eredményez stabilabb és gyorsabb konvergenciát, különösen mély neurális hálózatok esetében [19, 20].

1. Algoritmus SGD momentum variáns optimalizáló.

- 1: **Bemenet:** tanulási ráta α , momentum tényező γ
 - 2: **Inicializálás:** paraméterek $\theta \leftarrow \theta_0$, sebesség vektor $v \leftarrow 0$
 - 3: **while** a konvergencia feltételei nincsenek teljesítve **do**
 - 4: m elemű mintavétel a mintákból x_1, \dots, x_m a megfelelő címkékkel y_i .
 - 5: Számítsa ki a gradiens vektort $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x_i; \theta), y_i)$.
 - 6: Frissítse a sebességvektort: $v \leftarrow \gamma v - \alpha g$.
 - 7: Frissítse a paramétereket: $\theta \leftarrow \theta + v$.
 - 8: **end while**
-

2.3.2. Adam

Adam egy népszerű optimalizáló algoritmus (2. algoritmus), különösen hasznos a nagy és összetett neurális hálózatokban. Az algoritmus előnye, hogy adaptívan állítja be a tanulási rátát minden egyes paraméterhez, figyelembe véve a gradiensek első és második momentumját (átlagát és varianciáját). Bár az Adam-nak vannak hiperparaméterei (mint pl. a tanulási ráta, első és második momentumok leértékelési tényezői), a tapasztalatok szerint ezek a paraméterek kevésbé kritikusak, és az alapértelmezett beállítások általában jól működnek a különböző feladatokon[21, 20]. Az alábbi algoritmusban a g_t^2 a ún. Hadamard-szorzat, más néven elemi szorzat, két azonos méretű mátrix elem szerinti szorzása, ahol minden elem helyileg szorzódik meg. Ez a művelet a lineáris algebra és a gépi tanulás területén hasznos, különösen a neurális hálózatokban és a különféle mátrix műveletek során.

2. Algoritmus Adam optimalizáló.

- 1: **Bemenet:** α (tanulási ráta).
 - 2: **Bemenet:** β_1, β_2 (momentum és skála dekompozíciós tényezők, tipikusan 0.9 és 0.999).
 - 3: **Bemenet:** ϵ (numerikus stabilitási konstans, tipikusan 10^{-8}).
 - 4: **while** a konvergencia feltételei nincsenek teljesítve **do**
 - 5: $t \leftarrow t + 1$
 - 6: m elemű mintavétel a mintákból x_1, \dots, x_m a megfelelő címkékkel y_i .
 - 7: Számítsa ki a gradiens vektort $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x_i; \theta), y_i)$.
 - 8: Frissítse az első momentum vektort: $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$
 - 9: Frissítse a második momentum vektort: $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$
 - 10: Korrigálja az első momentum vektort: $\hat{m}_t \leftarrow \frac{m_t}{1 - \beta_1^t}$
 - 11: Korrigálja a második momentum vektort: $\hat{v}_t \leftarrow \frac{v_t}{1 - \beta_2^t}$
 - 12: Frissítse a paramétereket: $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$
 - 13: **end while**
-

2.3.3. RMSProp

Az RMSProp algoritmus (3. algoritmus) az adaptív tanulási ráta technikáinak egyike, amely a gradiensek négyzetének mozgó átlagát használja a paraméterek frissítéséhez, javítva ezzel a konvergencia stabilitását és sebességét [20].

3. Algoritmus RMSprop optimalizáló.

- 1: **Bemenet:** α (tanulási ráta)
 - 2: **Bemenet:** β (leértékelési tényező, tipikusan 0.9)
 - 3: **Bemenet:** ϵ (numerikus stabilitási konstans, tipikusan 10^{-8})
 - 4: Initializálás: $S \leftarrow 0$
 - 5: **while** a konvergencia feltételei nincsenek teljesítve **do**
 - 6: m elemű mintavétel a mintákból x_1, \dots, x_m a megfelelő címkékkel y_i .
 - 7: Számítsa ki a gradiens vektort $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x_i; \theta), y_i)$.
 - 8: Frissítse a gradiens négyzetek kumulatív összegét: $S \leftarrow \beta S + (1 - \beta)g^2$
 - 9: Frissítse a paramétereket: $\theta \leftarrow \theta - \alpha \frac{g}{\sqrt{S + \epsilon}}$
 - 10: **end while**
-

2.3.4. Keresztvalidáció

A keresztvalidáció egy statisztikai technika, amelyet a modellek teljesítményének objektív értékelésére használnak. Az adathalmazt több egyenlő részre osztják (foldokra) egy algoritmus segítségével (4. algoritmus), majd a modellt többször betanítják úgy, hogy minden iterációban egy rész kivételével az összeset használják a tanításhoz, és az utolsót a teszteléshez. Ezt a folyamatot annyiszor ismétlik meg, ahány részre osztották az adatokat. Az egyes iterációk eredményeit átlagolva kapunk megbízhatóbb becslést a modell általánosítási képességéről.

4. Algoritmus K-fold Keresztvalidáció.

- 1: **Bemenet:** Adathalmaz $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, ahol N az adatpontok száma
 - 2: **Bemenet:** K , a foldok száma
 - 3: **Felosztás:** Ossa fel az \mathcal{D} adathalmazt K diszjunkt részhalmazra (foldok), mindegyik körülbelül $\frac{N}{K}$ adatpontot tartalmaz
 - 4: **for** $k = 1$ to K **do**
 - 5: Jelölje \mathcal{D}_k a k -adik fold-ot, teszhalmazként
 - 6: Jelölje $\mathcal{T}_k = \mathcal{D} \setminus \mathcal{D}_k$, tanító halmazként
 - 7: **Modellezés:** Tanítsa be a modellt a \mathcal{T}_k tanító halmazon
 - 8: **Értékelés:** Mérje meg a modell teljesítményét a \mathcal{D}_k teszhalmazon, használva egy előre meghatározott metrikát
 - 9: **end for**
 - 10: **Aggregálás:** Számítsa ki az egyes foldok teljesítményméreteinek átlagát vagy mediánját.
 - 11: **Kimenet:** A modell átlagos teljesítménye a K fold-on keresztüli keresztvalidáció alapján.
-

3. fejezet

Többcímkes osztályozó építése

Ebben a fejezetben egy többcímkes osztályozó rendszert építünk fel az alapoktól. Az egyik leggyakrabban alkalmazott megközelítést használjuk: a konvolúciós neurális hálózatokat (CNN-eket).

A konvolúciós neurális hálók speciális mélytanulási modellek, amelyek különösen hatékonyak a képfeldolgozás terén. A CNN-ek automatikusan megtanulják és kinyerik a képek legfontosabb jellemzőit, így nincs szükség manuális képi jellemző kinyerésre. A rétegek hierarchikus felépítése lehetővé teszi, hogy az egyszerűbb mintázatoktól kezdve egyre összetettebb jellemzőket ismerjenek fel, mint például éleket, textúrákat és komplex alakzatokat. Ezáltal a CNN-ek különösen jól alkalmazhatók orvosi képfeldolgozási feladatokban, mint például a mellkasröntgen felvételek elemzése.

3.1. Konvolúciós neurális hálózatok

A CNN egy több rétegből álló mélytanulási modell. A rétegei közötti kapcsolatok a bemeneti képből kinyert jellemzők fokozatos finomításán és egyre összetettebb minták felismerésén alapulnak. A CNN-ek három fő rétegtípusból épülnek fel:

Konvolúciós réteg: Ez a réteg kis, mozgó kerneleket használ, hogy lokális jellemzőket nyerjen ki a bemeneti képből. A szűrők a kép különböző régióiban azonos mintákat keresnek, például éleket és vonalakat.

5. Definíció. Legyen I egy szürkeskálás digitális kép, amely egy $W \times H$ méretű rácson értelmezett függvény. A konvolúciós réteg egy vagy több kernelt alkalmaz a képre, amelyek mindegyike egy $K \times K$ méretű mátrix. A kernelek a bemeneti képen végighaladva (konvolválva) létrehoznak egy kimeneti jellemző térképet az alábbi módon:

$$Y(i, j) = (I * K)(i, j) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} I(i + m, j + n) \cdot K(m, n) + b,$$

ahol $0 \leq i \leq W - K$ és $0 \leq j \leq H - K$. $I(i + m, j + n)$ a bemeneti kép pixelértéke az $(i + m, j + n)$ pozícióban, $K(m, n)$ a szűrő értéke az (m, n) pozícióban, és b egy bias érték. Az $I * K$ jelöli a konvolúciós műveletet.

Pooling réteg: Ez a pooling réteg az előző rétegből származó jellemzők

méretét csökkenti, miközben megőrzi a legfontosabb információkat. A leggyakrabban használt verziója a maximum pooling. A képből kiválasztják a legnagyobb értékű jellemzőket egy adott ablakon belül, ezzel csökkentve a kép dimenzióit és növelve a modell robusztusságát. A pooling réteg használja a *stride* paramétert, amely meghatározza, hogy az ablak milyen lépésközönként mozog végig a képen. Ez a paraméter tovább csökkenti a kimeneti tér méretét és növeli a számítási hatékonyságot.

6. Definíció. Legyen Y egy konvolúciós réteg kimeneti jellemző térképe, amely egy $W' \times H'$ méretű rácson értelmezett függvény. Az összevonó réteg célja a jellemző térkép méretének csökkentése, miközben megőrzi a legfontosabb jellemzőket. Az egyik leggyakrabban használt összevonó művelet a maximum összevonás, amely az alábbi módon definiálható:

$$Y_p(i, j) = \max_{0 \leq m < P, 0 \leq n < P} Y(i + m, j + n),$$

ahol $P \times P$ az összevonó ablak mérete, és $Y_p(i, j)$ az összevonó réteg kimeneti értéke az (i, j) pozícióban. Az összevonás művelet során a $P \times P$ ablak minden egyes pozíciójában a legnagyobb értéket választja ki, így csökkentve a jellemző térkép dimenzióit.

Teljesen összekapcsolt réteg: A feladatuk, hogy a korábbi rétegekből származó jellemzőket egy vektorba lapítsák (flatten), majd egy hagyományos neurális hálót (dense rétegek) alkalmazzanak az osztályozásra. Ezek a ré-

tegek a bemeneti jellemzők és a kimeneti osztályok közötti nemlineáris összefüggéseket tanulják meg, és hozzájárulnak a modell általános teljesítményéhez. Jelen dolgozatban a CNN-ek után lévő dense réteg(ek) megnevezésére használjuk.

A CNN rétegei által kinyert jellemzők fokozatosan egyre összetettebbé válnak: az első konvolúciós rétegek az egyszerű képi jellemzőket, mint például éleket és vonalakat, a következő rétegek pedig bonyolultabb mintákat, alakzatokat és textúrákat ismernek fel. A legfelső rétegek a tárgyak magas szintű reprezentációit nyerik ki, amelyeket a végső osztályozási döntések meghozatalára használnak.

3.2. Regularizációs technikák

A mélytanulási modellek teljesítményének javítása és a túlilleszkedés elkerülése érdekében számos regularizációs technika alkalmazható. Ezek a módszerek segítenek stabilabb és robosztusabb modellek kialakításában. Ebben az alfejezetben három fontos regularizációs technikát tárgyalunk amelyeket használtunk jelen dolgozat során: a dropout, az early stopping és a reduce on plateau. Mindhárom módszer különböző megközelítést alkalmaz a modell tanulási folyamatának szabályozására és optimalizálására, míg az első a modellbe beéptett regularizáció addig a másik kettő pedig a tanítási folyamatba avatkozik be a túltanulás elkerülése érdekében.

Dropout: A dropout egy hatékony regularizációs technika, amelyet a mély neurális hálózatok túlilleszkedésének csökkentésére használnak. Ennek a módszernek a lényege, hogy az tanítás során véletlenszerűen, egy előre meghatározott valószínűséggel "kihagyja" a neurális hálózat egyes neuronjait. Ez megakadályozza, hogy a neurális hálózat túlságosan függjön egy adott résztől, és arra kényszeríti a hálózatot, hogy a tanulás során jobban általánosítson. A dropout alkalmazása jelentősen javíthatja a modell általánosítási képességét és stabilitását.

Early Stopping: Az early stopping egy olyan regularizációs technika, amely közvetlenül a tanítási folyamatba avatkozik be annak érdekében, hogy elkerülje a modell túlilleszkedését. A módszer lényege, hogy a tanítási folyamatot akkor állítja meg, amikor a modell teljesítménye a validációs adathalmazon már nem javul tovább. Ezzel megelőzhető, hogy a modell a tréning adatokhoz túlzottan igazodjon, és ezáltal rosszul általánosítson új, ismeretlen adatokon. Az early stopping segítségével a modell optimális állapotban tartható, biztosítva a legjobb teljesítményt és általánosítást.

Reduce on Plateau: A reduce on plateau egy adaptív tanulási ráta szabályozó technika, amely a tanítási folyamat során csökkenti a tanulási rátát, ha a modell teljesítménye egy bizonyos időn keresztül nem javul. Ennek a módszernek az a célja, hogy a tanulási folyamat során finomhangolja a tanulási rátát, elkerülve ezzel a túl gyors konvergenciát és a helyi minimumokban való megrekedést. A technika hatékonyan kombinálható más

regularizációs módszerekkel, mint például az early stopping, a stabilabb és általánosíthatóbb modellek érdekében.

3.3. Javasolt architektúra

A javasolt módszerben először a CNN blokkot kell meghatároznunk (ld. 3.1. táblázat). Bemenetként pedig eredeti, nem módosított röntgenképeket használ. A blokk 2 darab azonos méretű konvolúciós réteggel rendelkezik. A konvolúciós rétegek után normalizáljuk az értékeket majd egy maximum pooling réteggel csökkentjük a dimenziókat. Ez a megközelítés VGG [24]-szerű architektúra, mivel 3×3 kernelt használ egymásra halmozva, és blokkonként maximum pooling réteget használ a méret csökkentésére.

3.1. táblázat. Általunk javasolt konvolúciós blokk bemutatása.

| Megnevezés | Méret |
|-------------------|-----------------------|
| Konvolúciós réteg | $u \times 3 \times 3$ |
| Konvolúciós réteg | $u \times 3 \times 3$ |
| Normalizáció | |
| Maximum Pooling | 2×2 |

A javasolt módszereink osztályozó részében két dense réteg van 1024 egységgel közte pedig dropout réteg $p = 0,5$ értékkel (ld. 3.2. táblázat), és az utolsó réteg pedig 14 kimeneti egységgel rendelkezik. Az a

2.2.6. szekcióban bemutatott szigmoid (2.4) aktivációs függvényt használja. Amint látható a három javasolt osztályozó teljesen összekapcsolt rétegei megegyeznek és csak a konvolúciós részben van különbség, mivel a Modell A kevesebb konkolúciós blokkot tartalmaz így a kinyert tulajdonságvektor számottevően nagyobb mint a másik két modell esetén, ezért teljesen összekapcsolt réteg megnövekedett paraméter száma miatt az összes tanítható paraméterek száma többszöröse lesz mint a Modell B és C esetén.

Ebben a módszerben az Adam (ld. 2. algoritmus) optimalizálót használtuk, mely a leggyakoribb optimalizációs algoritmus, ha egy üres modellt szeretnénk tanítani, az optimalizáló paramétereit az alapértelmezett beállításokon hagytuk $\beta_1 = 0,9$ és $\beta_2 = 0,999$. Veszteségfüggvényként a 2.2.6. szekcióban definiált bináris keresztentropia többcímkes módosított verzióját használjuk (2.8).

3.4. Eredmények

A kísérleti eredmények összehasonlítása három különböző modell között történt, amelyek különböző architektúrákat használnak a mellkasröntgen felvételek többcímkes osztályozására. Az eredményeket az 3.3. táblázatban foglaltuk össze. Az összehasonlítás során a Modell A, Modell B és Modell C teljesítményét értékeltük különböző tüdőbetegségek detektálásában. A tesztek 50-es epoch számmal futottak és 3 stagnáló validáló AUC

3.2. táblázat. Javasolt architektúra.

| | Modell A | Modell B | Modell C |
|------------------------------|----------|----------|----------|
| CNN Blokk (kernel méret) | 32 | 32 | 32 |
| CNN Blokk | 64 | 64 | 64 |
| CNN Blokk | 64 | 64 | 64 |
| CNN Blokk | 128 | 128 | 64 |
| CNN Blokk | 128 | 128 | 128 |
| CNN Blokk | - | 128 | 128 |
| CNN Blokk | - | - | 128 |
| CNN Blokk | - | - | 128 |
| Flatten | | | |
| Dense (unit szám) | 1024 | 1024 | 1024 |
| Dropout | 0,5 | 0,5 | 0,5 |
| Dense | 1024 | 1024 | 1024 |
| Dropout | 0,5 | 0,5 | 0,5 |
| Dense | 14 | 14 | 14 |
| Hálózat paramétereinek száma | 27338670 | 8366766 | 2445102 |

esetén 10-edére csökkentettük a tanulási rátát, 6 stagnáló validációs AUC esetén pedig early stopping-ot alkalmaztunk.

Mind a három modell viszonylag egyszerű CNN architektúra, változó mélység és kernel darabszámmal. A Modell A négy CNN blokkot használ, minden blokkban növekvő számú kernel számmal (32, 64, 128, 128). A Modell B ugyanezen struktúrán alapul, de egy további CNN blokkot tartalmaz, ami 128 kernelt használ. A Modell C a legkomplexebb, összesen hét CNN blokkot használva, ahol az utolsó négy blokk mindegyike 128 kernelt alkalmaz. Mindhárom modell esetében a CNN blokkokat teljesen összekapcsolt rétegek (dense rétegek) követik, amelyek 1024 neuront tartalmaznak, dropout rétegekkel a túlillesztés csökkentése érdekében, végül egy 14 neuront tartalmazó kimeneti réteg következik, amely a különböző betegségek jelenlétét jelzi.

Az eredmények azt mutatják, hogy a Modell C általában jobb teljesítményt nyújtott, mint a Modell A és Modell B, különösen a kardiomegália, térfoglalás és légmell címkékben. Az átlagos pontosság is a Modell C esetében volt a legmagasabb ($0,758 \pm 0,017$), míg a Modell B közepes teljesítményt nyújtott ($0,741 \pm 0,013$), és a Modell A a legkevésbé volt pontos ($0,696 \pm 0,031$). Ez az eredmény azt mutatja, hogy a nagyobb mélységű és több CNN blokkot tartalmazó modellek általában jobban teljesítenek a többcímkes osztályozási feladatokban, bár a számítási erőforrások szükségére jelentősen növekedett.

3.3. táblázat. Eredmények (AUC) a ChestX-ray14 adathalmazon.

| | Modell A | Modell B | Modell C |
|--------------------|-----------------|-----------------|-----------------|
| Atelectasis | 0,722 ± 0,025 | 0,754 ±0,000 | 0,770 ±0,020 |
| Cardiomegaly | 0,755 ± 0,056 | 0,817 ±0,029 | 0,867 ±0,012 |
| Consolidation | 0,746 ± 0,023 | 0,771 ±0,014 | 0,785 ±0,012 |
| Edema | 0,833 ± 0,021 | 0,871 ±0,010 | 0,872 ±0,018 |
| Effusion | 0,792 ± 0,043 | 0,845 ±0,013 | 0,862 ±0,013 |
| Emphysema | 0,681 ± 0,044 | 0,740 ±0,026 | 0,750 ±0,029 |
| Fibrosis | 0,682 ± 0,015 | 0,724 ±0,009 | 0,715 ±0,017 |
| Hernia | 0,620 ± 0,021 | 0,655 ±0,065 | 0,678 ±0,044 |
| Infiltration | 0,641 ± 0,013 | 0,656 ±0,002 | 0,665 ±0,005 |
| Mass | 0,666 ± 0,051 | 0,729 ±0,017 | 0,764 ±0,045 |
| Nodule | 0,591 ± 0,037 | 0,639 ±0,015 | 0,647 ±0,030 |
| Pleural Thickening | 0,672 ± 0,040 | 0,720 ±0,009 | 0,741 ±0,025 |
| Pneumonia | 0,672 ± 0,014 | 0,690 ±0,004 | 0,711 ±0,028 |
| Pneumothorax | 0,676 ± 0,060 | 0,759 ±0,015 | 0,791 ±0,013 |
| Átlag | 0,696± 0,031 | 0,741 ± 0,013 | 0,758 ± 0,017 |

3.5. Összefoglalás

Több radiológust nem tudunk előállítani, de a bemutatott módszer kiváló alapot nyújt egy teljes körűen automatizált mellkasröntgen-felvétel jelölő eszköz kifejlesztéséhez a jövőbeni munkák során. Az eddigi eredmények alapján megállapítottuk, hogy a három különböző modell (Modell A, Modell B és Modell C) teljesítménye a mellkasröntgen felvételek többcímkes osztályozásában nem érte el a kívánt 0,76 AUC-értéket. A legjobb átlagos AUC-t a Modell C érte el, amely $0,758 \pm 0,017$ volt, azonban ez sem haladta meg a 0,76 küszöböt. Az eredmények arra utalnak, hogy bár a Modell C

több CNN blokkot tartalmaz, és jobb teljesítményt nyújt, még mindig van lehetőség a pontosság növelésére.

Ezeket az eredményeket figyelembe véve, a továbbiakban ismert és bevált CNN architektúrákat fogunk adaptálni és finomhangolni a többcímű röntgenfelvételek osztályozására. Olyan architektúrákat, mint például a ResNet, Inception és DenseNet, fogunk használni alapként, és különböző technikákkal, mint a transzfer tanulás és a hiperparaméter-optimalizálás, végezetül pedig ensemble technikákkal, próbáljuk meg növelni az AUC értékét és javítani az osztályozási teljesítményt.

Az a célunk, hogy olyan modelleket hozzunk létre, amelyek képesek pontosabb és megbízhatóbb diagnosztikai eredményeket nyújtani a mellkasröntgen felvételek elemzése során.

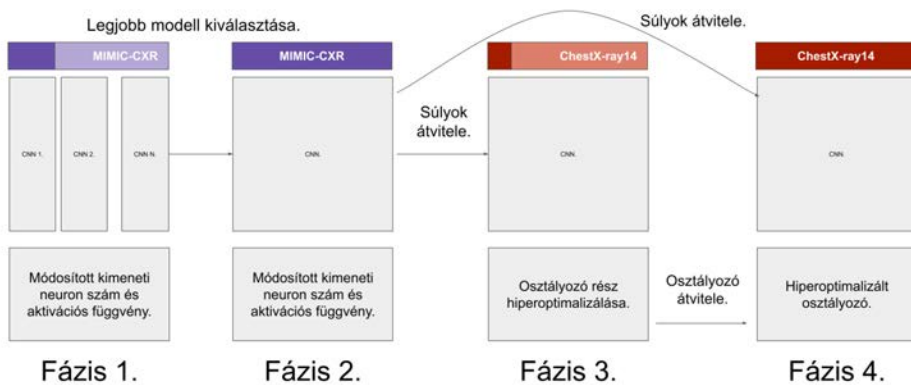
4. fejezet

Hiperparaméter-optimalizált osztályozó létrehozása többcímkes osztályozásra

Jelen fejezetben transzfer tanulást kombináljuk hiperoptimalizációs technikával és ennek segítségével adunk meg egy új és jobb módszert a többcímkes röntgenfelvételek osztályozására. Ehhez jól ismert konvolúciós neurális hálózatokra, hiperoptimalizálás segítségével építünk új osztályozó részt. Ezt folyamatot négy fázisra bontottuk, az átláthatóság érdekében. Az első fázisban, néhány jól ismert architektúra gyors tesztelését végeztük el a MIMIC-CXR adatkészlet részhalmazán. Az első lépés célja, hogy az adott CNN architektúra alkalmas-e többcímkes osztályozásra röntgenké-

pek esetén. A használt jól ismert architektúrák (VGG, ResNet, DenseNet) alapvetően kategorikus osztályozásra lettek kifejlesztve az ImageNet adatbázison. Ezen periódusban az ImageNet-en betanított súlyokkal inicializált modellt tanítjuk tovább, így hamar láthatóvá válik, hogy képes-e tanulni.

Ezt követő fázisban kiválasztottuk a legjobb konfigurációt, amit a teljes MIMIC-CXR adatkészleten tanítottunk, ezzel elkészítve a saját előtanított súlyainkat a további fázisokhoz. A 3. fázisban már az összehasonlíthatósághoz kiválasztott ChestX-ray14 adatbázist használjuk. Ebben a lépésben teljesen eltávolítjuk a nyertes CNN architektúra osztályozó részét és hiperoptimalizáló algoritmus segítségével új osztályozó részt építünk rá.



4.1. ábra. Folyamatábra a többlépcsős optimalizációs módszerünkről.

Az utolsó, 4. fázisban az optimális hiperparaméterek használatával elvégeztük az ajánlott modell tanítását a hivatalos adatkészlet szeleteken, hogy összehasonlíthatóvá váljon a módszerünk a legkorszerűbb algoritmu-

sokkal. Végezetül a végleges modell javulását Grad-Cam módszer segítségével vizualizáljuk.

4.1. Irodalmi áttekintés

A képfeldolgozási módszerek fejlődésével új megközelítések láttak napvilágot. Ennek egyik kiemelkedő ága a konvolúciós neurális hálózatok (CNN-ek). Ezek speciális típusú mély neurális hálózatok, amelyeket elsősorban képi adatok feldolgozására terveztek. A CNN-ek kiemelkedően hatékonyak különböző látással kapcsolatos feladatokban, mint például az arc- és objektumfelismerés, képszegmentálás és képkategorizálás. Az architektúra úgy lett optimalizálva, hogy kiaknázza a képek térbeli hierarchiáját, és képes legyen azonosítani és klasszifikálni a vizuális mintákat magas szintű pontossággal.

Tovább lépve a bevezetőben említett megoldásokon (ld. 1. fejezet), természetesen ezek a mélytanulási technikák is alkalmazhatók a képi jellemző szinopszis automatizálására, de ehhez nagyszámú annotált képre és szignifikánsan több hardveres erőforrásokra van szükség, mint a hagyományos kézzel kinyert esetben. A grafikus kártyák árának drasztikus csökkenése enyhítette az erőforrásproblémákat. Ezenkívül számos nagymintás mellkasröntgen felvétel adatbázis vált elérhetővé. A jelen fejezet egy mélytanulási alternatívára összpontosít, automatizált jellemző-kivonással. Több

publikáció is rendelkezésre áll, amelyek a többcímkes mellkasröntgen felvételek osztályozásával foglalkoznak: Wang és mtsai. [5] klasszikus mélytanulási architektúrák mellkasröntgen felvételekhez való használatát írták le. Egy másik kutatócsoport, Rajpurkar és mtsai. [12] a jól ismert architektúrát, a DenseNet121-et használta a jellemző kivonatoláshoz, és csak az utolsó teljesen összekapcsolt réteget változtatta meg, hogy alkalmazkodjon a többcímkes osztályozás sajátosságaihoz. Néhányan további adatokat használnak a címkézéshez; Wang és mtsai. [31] külső radiológiai jelentéseket használtak a teljesítmény javítása érdekében, Tang és mtsai. [32] pedig a radiológiai jelentésekből kinyert súlyossági szintek segítségével azonosították a betegségkategóriákat. Yao és mtsai. [33] szintén a DenseNet-et használták; ők azonban kódolóként használták, és dekódolóként a LSTM kombinálták. Guan és mtsai. [34] a ResNet-et használták gerincként, és kategóriánként attention learning-et valósítottak meg. Wang és mtsai. [35] szintén a DenseNet 121-et használták kutatásukhoz, hogy komplex hármas attention learning-et hozzanak létre a problémához; a három modul a csatorna-, a címke- és a méretalapú figyelem tanulás.

A jelenlegi legmodernebb architektúrák átfogó áttekintése azt mutatja, hogy a jól ismert CNN-ek jellemzői extrakciós rétegeinek módosítására összpontosítanak. Bár ezek a megközelítések természetesen előrelépést jelentenek a többcímkes osztályozási területen, a szakirodalom áttekintése azt mutatja, hogy a hálózatok osztályozó komponensének változtatásai egy

alulkutatott terület, ezért mi most ebben mélyedünk el.

4.2. Módszertan

4.2.1. Konvolúciós neurális hálózatok

A CNN-ek a mélytanulás egyik legfontosabb eszközei, különösen a képfeldolgozási feladatok terén. Ezek a hálózatok speciálisan arra lettek tervezve, hogy felismerjenek és automatikusan nyerjenek ki jellemzőket a képekből, ezáltal lehetővé téve a gépi tanulási modellek számára, hogy hatékonyabban és pontosabban dolgozzanak fel vizuális adatokat. A CNN architektúra több rétegből áll, amelyek közül a legfontosabbak a konvolúciós rétegek, amelyek a képi adatokból szűrő kernelek segítségével jellemzőket vonnak ki. Ezek a szűrők képesek különböző vizuális minták, mint például élek, szögek vagy textúrák felismerésére.

A CNN-ek hatékonysága részben az ún. súlymegosztási mechanizmusnak köszönhető, ami azt jelenti, hogy a hálózat ugyanazokat a szűrőket használja a kép különböző részein. Ez nemcsak csökkenti a modell paramétereinek számát, hanem lehetővé teszi a hálózat számára, hogy invariánsak legyenek a képen belüli objektumok helyzetével szemben. Továbbá, az összevonó (pooling) rétegek csökkentik az adatok dimenzióját, miközben megőrzik a fontos jellemzőket, ezáltal növelve a hálózat hatékonyságát és csökkentve a túltanulás (overfitting) esélyét.

A CNN-ek alkalmazási területe rendkívül széleskörű, a képfelismeréstől és osztályozástól kezdve az orvosi képfeldolgozáson át a videóanalízisig és a természetes nyelvfeldolgozásig (NLP) is, ahol a szöveges adatokat "képként" kezelik. Különösen az orvosi képfeldolgozásban mutatkozott meg a CNN-ek forradalmi potenciálja. Ahol képesek diagnózisok támogatására, anomáliák azonosítására, és akár olyan részletességgel, ami az emberi szakértők számára is kihívást jelenthet. Az egészségügyi diagnosztika mellett a CNN-eket használják járművek automatikus vezetési rendszereiben, biztonsági megfigyelésben és sok más területen, ahol a vizuális adatok értelmezése kulcsfontosságú.

VGG architektúra

Első pillantásra a VGG egy egyszerű architektúrának tűnik, mégis fontos előrelépés volt a számítógépes látás fejlődésében. Ez egy CNN-modell, amelyet a K. Simonyan és A. Zisserman [24] kutatópáros fejlesztett ki. A VGG elnevezés a tanszéktől (Visual Geometry Group) származik, ahol kifejlesztették. Ennek az architektúrának kétféle konfigurációja van, a VGG16 és 19; a névben szereplő számok a hálózatok mélységére utalnak. Ez a két hálózat kis paraméterekkel rendelkezik a teljesen összekapcsolt blokk nélkül (az eredeti beállításban két ilyen réteggel rendelkeznek, egyenként 4096 egységgel), így nagy szabadságot kapunk ezen osztályozó rétegek kialakításában. A jelen munkában a 16 mélységű konfigurációt használtunk.

ResNet architektúra

A ResNet, a Residual Network rövidítése, egy úttörő neurális hálózati architektúra a mélytanulás területén. A ResNet előtt a hagyományos neurális hálózatoknak volt egy korlátja, amely azt foglalja magában, hogy minél több rétegű volt annál nehezebb volt tanulniuk. Ez főként az úgynevezett "eltűnő gradiensek" nevű problémának volt köszönhető, amikor a hálózat képzéséhez használt jelek sok rétege alatt olyan kicsivé váltak, hogy a tanulás ténylegesen leállt. Ennek leküzdésére a ResNet egy újszerű koncepciót vezetett be. Ahelyett, hogy minden egyes rétegben teljesen új reprezentációkat tanult volna, az előző réteg különbségein, így a "maradékoknak" a tanulására összpontosított. Ezt a megközelítést a Highway-hálózatok és a hosszú rövid távú memória (LSTM) hálózatok elvei ihlették, amelyek szintén arra törekszenek, hogy az információt sok rétegen keresztül megőrizzék[36]. A ResNet architektúráját a "rövidített kapcsolatok" jellemzik. Ezek a rövidítések lehetővé teszik, hogy a hálózat átgorjon néhány réteget, és az információt közvetlenül a későbbi rétegeknek adja át. Ez nem csak az eltűnő gradiens probléma leküzdésében segít, hanem lehetővé teszi a több száz rétegből álló hálózatok sikeres képzését is, amik sokkal mélyebbek, mint amik korábban lehetségesek voltak. A ResNet egyik népszerű változata a ResNet50V2. A hálózatnak ebben a változatában az 50 réteg öt blokkba van szervezve, amelyek mindegyike több réteget tartalmaz ezen gyorskapcsolatokkal. Ez a struktúra tette a ResNet50-t széles körben

használhatóvá különböző alkalmazásokban, a képfelismeréstől a fejlett számítógépes látási feladatokig. Jelen dolgozatban a teljesítmény és memória optimalizált V2 verzióját használjuk.

DenseNet architektúra

Ez a fajta hálózat hasonlít a reziduális neurális hálózathoz, néhány különbséggel. A DenseNet a kimenet-bemeneti dimenziók összehangolásával a hálózat összes kimenetét bemenetként használja a hálózat összes fenti rétegéhez [37]. Mindegyik konfiguráció négy egyforma blokkot tartalmaz; az egyetlen különbség ezeken a blokkokon belül a konvolúciós rétegek száma. Az összes blokk után van egy transzformációs blokk, amely a dimenzió csökkentést végzi. A konvolúciós blokkok száma a következő: DenseNet-121 (6, 12, 24, 16), DenseNet-169 (6, 12, 32, 32, 32) és DenseNet-201 (6, 12, 48, 32). Ebben a tanulmányban a DenseNet 121-es verzióját használjuk.

4.2.2. Sigmoid veszteségfüggvény

A bevezetőben (ld. 2.2.6 szekció) bemutatott többcímű keresztentropia mellett ebben a fejezetben egy másik veszteségfüggvényt is használunk. Ez pedig a szigmoid fókuszú veszteségfüggvény. Ez a függvény (4.1) egyenlet a standard bináris keresztentropia veszteség adaptációja. Különösen hasznos az adathalmazok osztály egyenlőtlenségének kezelésére. A szigmoid fókuszú veszteségfüggvényt a következőképpen határozzuk meg

$$L_{focal} = - \sum_{j=1}^S \sum_{i=1}^N \alpha (1 - \hat{y}_{j,i})^\gamma \log(\hat{y}_{j,i}), \quad (4.1)$$

ahol $\hat{y}_{j,i}$ a modell becsült valószínűsége az j -edik kép, i -edik címkéjén, az $y = 1$ címkével rendelkező osztályra (S a képek száma, N a címkék száma), α pedig az osztály súlyozási tényezője, amelyet általában az osztályok kiegyensúlyozására állítunk be. A γ a fókuszáló paraméter, amely a könnyű példák lefelé súlyozásának mértékét állítja be. A függvény dinamikusan skálázza a keresztentropia veszteséget az előrejelzések megbízhatósága alapján. Csökkenti a veszteséget a jól osztályozott példák esetében, és jobban összpontosít a nehezen tévesen osztályozott példákra.

4.2.3. Hiperparaméter optimalizáció

A 3. fázisban, ahogy az 4.1. ábrán látható, hiperparaméter-optimalizálási módszereket alkalmaztunk, hogy az előzetesen betanított CNN fölé egy új osztályozó részt építsünk. A hiperparaméter-hangolás vagy -optimalizálás (néha HPO) egy olyan módszer, amellyel a potenciálisan optimális paraméterkészletből meghatározhatók az optimális paraméterek. A gépi tanulás széles körben használt technika a tanulási folyamathoz kapcsolódó paraméterek, például a batch mérettel, a tanulási sebesség stb. optimalizálására. Ezzel szemben ebben a tanulmányban néhány jól ismert architektúra esetében az egész, teljesen összekapcsolt réteg felépítésére próbáltuk használni.

Jelen munkában a Keras Tuner [38] szoftverkönyvtárat használtuk.

Módszertanunk egyik kulcsfontosságú eleme a Hyperband algoritmus, egy hatékony, többféle hűséggel működő optimalizálási módszer beépítése a modellünk képzési folyamatába. A Hyperband a szukcesszív felező algoritmusok hatékony kiterjesztéseként működik, amelyek koncepcionálisan egy előre meghatározott számítási költségvetés hatékony felhasználásában gyökereznek [39].

Tekintettel a mellkasröntgen felvételek címkézésének többcímkes problémájának összetettségére és nagyságrendjére, a Hyperband azon képessége, hogy többféle konfigurációt képes egyidejűleg feldolgozni és gyorsan kiválasztani az ígéretes konfigurációkat, alapvető fontosságú. Ez a megközelítés lehetővé teszi, hogy hatékonyan navigáljunk a kiterjedt hiperparaméter térben, és a párhuzamos feldolgozási képesség drámaian felgyorsítja az optimalizálási folyamatot. Ezek az előnyök teszik a Hyperbandet megfelelő választássá erre a feladatra.

Számos más hiperoptimalizálási keretrendszer is rendelkezésre áll. Az egyik legjelentősebb az Optuna [40]. Az Optuna nagyobb rugalmasságot és funkciókat kínál, mint a Keras Tuner, például metszést és felhasználóbarát felületet, ami előnyösebbé teheti olyan forgatókönyvek esetében, amelyben Hyperband algoritmikus hatékonysága kevésbé kritikus, mint más tényezők.

4.3. Eredmények

Ennek a fejezetnek a célja a mellkasi röntgenfelvételek többcímkes osztályozásának optimalizálása volt, fejlett mélytanulási technikák alkalmazásával. Kutatásunk két robusztus és független adathalmaz, a MIMIC-CXR és a ChestX-ray14 használatára összpontosít, nem képi jellemzők vagy egyéb hozzáadott információk nélkül.

4.3.1. Irodalmi architektúrák vizsgálta

Az első fázisban (ld. 4.1. ábra) ismert CNN architektúrák gyors vizsgálatát végeztük, hogy képesek-e helytállni többcímkes környezetben röntgenfelvételek esetén. Három jól ismert architektúrát használtunk: DenseNet121, ResNet50V2 és VGG16. Mindegyik modell egyedi szerkezeti tulajdonságai és a képosztályozási feladatokban bizonyított hatékonysága miatt választottuk. Ezen modellekhez a hivatalos TensorFlow implementációt használtuk fel fejrész nélkül, a beépített ImageNet súlyokkal. A többcímkes adaptációhoz egy szigmoid aktivációs függvénnyel rendelkező teljesen összekapcsolt réteget adtunk hozzá 13 egységgel, hogy illeszkedjen a MIMIC-CXR adatbázis címkéihez. Mivel ImageNet-súlyok használatához háromcsatornás képekre van szükség, amihez módosítani kell a bementi képeknek a méretét 224×224 méretre. Először a szürke képet háromcsatornás képpé bővítettük, majd a MIMIC-CXR esetében 1377×1153 -ról

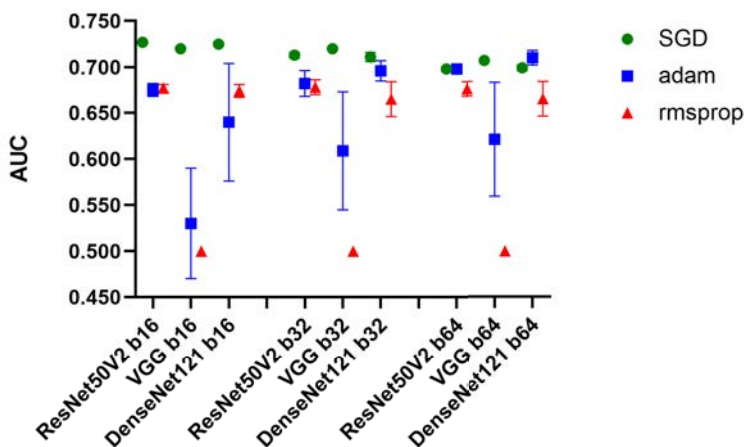
(a ChestX-ray14 esetében 1024×1024 -ról) 256×256 -ra átméreteztük a képet. S ezt követően a képeket méretre vágjuk (224×224), majd a tanító képek esetén ezen vágást véletlenszerű margók segítségével végezzük, amíg a tesztelő képeknél középre igazítva vágunk. A kép augmentációnál (ld. 2. definíció) véletlenszerű horizontális tükrözést és forgatásokat alkalmaztunk $-0,4$ és $0,4$ radián között. Folyamatunk első fázisban (1. ábra) a magas szintű gyorstesztek folytak, amelyeket nagyon kis mennyiségű MIMIC-CXR adat felhasználásával végeztünk el (tanítás: 8000, validálás: 1500, tesztelés: 2500). Az optimalizálás során különböző 16, 32 és 64 elemű kötegekkel és különböző optimalizálókkal kísérleteztünk, nevezetesen a bevezetésben bemutatott Adam (2. algoritmus), a SGD (1. algoritmus) és az RMSProp (3. algoritmus) (4.1. táblázat) segítségével. Ezek a paraméterek kulcsfontosságúak voltak a modellek finomhangolásában, mivel a kötegek elemszáma befolyásolja a modell általánosító képességét, míg az optimalizálók kulcsszerepet játszanak a tanulási folyamat sebességében és stabilitásában. A folyamat során 10 iterációt és 0,001-es tanulási rátát használtunk, mivel e fázis jelenlegi célja csak az volt, hogy magas szintű elemzést kapjunk az architektúrák tanulási képességeiről. Az optimalizáló algoritmusok esetében az alapkonfigurációkban alkalmaztuk, ez az Adam esetében $\beta_1 = 0,9$ és $\beta_2 = 0,999$, az RMSProp és az SGD esetében pedig $momentum = 0,0$ volt.

A DenseNet121, a ResNet50V2 és a VGG16 is konzisztens teljesít-

4.1. táblázat. A modellek teljesítménye az átfogó gyorseszten.

| Architektúra neve | Tanítható paraméterek | Köteg méret | Optimalizáló | AUC |
|-------------------|-----------------------|-------------|--------------|---------------|
| ResNet50V2 | 23,5M | 16 | SGD | 0,727 ± 0,005 |
| DenseNet121 | 6,9M | 16 | SGD | 0,726 ± 0,002 |
| VGG16 | 134,3M | 32 | SGD | 0,723 ± 0,001 |
| VGG16 | 134,3M | 16 | SGD | 0,720 ± 0,001 |
| ResNet50V2 | 23,5M | 32 | SGD | 0,713 ± 0,003 |
| DenseNet121 | 6,9M | 32 | SGD | 0,711 ± 0,005 |
| DenseNet121 | 6,9M | 64 | Adam | 0,710 ± 0,008 |
| ResNet50V2 | 23,5M | 64 | Adam | 0,709 ± 0,010 |
| VGG16 | 134,3M | 64 | SGD | 0,707 ± 0,001 |
| DenseNet121 | 6,9M | 64 | SGD | 0,699 ± 0,003 |
| ResNet50V2 | 23,5M | 64 | SGD | 0,698 ± 0,002 |
| DenseNet121 | 6,9M | 32 | Adam | 0,696 ± 0,011 |
| DenseNet121 | 6,9M | 64 | RMSprop | 0,688 ± 0,024 |
| ResNet50V2 | 23,5M | 32 | Adam | 0,682 ± 0,014 |
| ResNet50V2 | 23,5M | 32 | RMSprop | 0,678 ± 0,008 |
| ResNet50V2 | 23,5M | 16 | RMSprop | 0,677 ± 0,004 |
| ResNet50V2 | 23,5M | 64 | RMSprop | 0,676 ± 0,008 |
| ResNet50V2 | 23,5M | 16 | Adam | 0,675 ± 0,007 |
| DenseNet121 | 6,9M | 16 | RMSprop | 0,674 ± 0,007 |
| DenseNet121 | 6,9M | 32 | RMSprop | 0,665 ± 0,019 |
| DenseNet121 | 6,9M | 16 | Adam | 0,640 ± 0,067 |
| VGG16 | 134,3M | 64 | Adam | 0,621 ± 0,062 |
| VGG16 | 134,3M | 32 | Adam | 0,609 ± 0,064 |
| VGG16 | 134,3M | 16 | Adam | 0,530 ± 0,060 |
| VGG16 | 134,3M | 64 | RMSprop | 0,500 ± 0,000 |
| VGG16 | 134,3M | 16 | RMSprop | 0,500 ± 0,000 |
| VGG16 | 134,3M | 32 | RMSprop | 0,500 ± 0,000 |

ményt mutatott a különböző elemszámú kötegek között. Figyelemre méltó, hogy az AUC csökkenése minimális volt a kötegek elemszámának növekedésével, ami arra utal, hogy ezek a modellek még nagyobb kötegek feldolgozásakor is megőrzik hatékonyságukat. A 4.2. ábrán látható, hogy az SGD optimalizáló következetesen jobb volt a többinél.



4.2. ábra. Tanulási teszt a MIMIC-CXR adatkészleten.

A szórások általában alacsonyak voltak, ami azt jelzi, hogy a modellek különböző futtatásai során következetes teljesítményt mutattak. Az Adam optimalizálóval azonban a szórás jelentős növekedése volt megfigyelhető, különösen a VGG16 esetében, ami azt jelezheti, hogy ezekben a konfigurációkban kevésbé stabil a teljesítmény.

4.3.2. Röntgen specifikus súlyok létrehozása

A második fázisban, ahogy az 4.1. ábrán is látható, kizárólag a ResNet50V2-t használtuk a további kísérletekhez. Ezt a döntést több tényező együttesen befolyásolta, többek között a modell számítási hatékonysága és egyszerűsége. A módszertani keretrendszerünk következő sarokköve a transzfer tanulás. A transzfer tanulás során egy modellt egy nagy adathalmazon képezünk ki, majd egy másik általában kisebb adathalmazra adaptáljuk. Ez a megközelítés különösen előnyös az orvosi képalkotás területén, ahol ritkák a nagy, annotált adatkészletek. Az egyik feladatból szerzett tudás felhasználásával a transzfer tanulás javítja a tanulás hatékonyságát és a tudás általánosítását egy másik feladatra. Jelen szakaszban először a ResNet50V2-t a teljes MIMIC-CXR-adatkészleten tanítottuk, amely mellkasi röntgenfelvételek széles skáláját nyújtotta. Ezen adathalmazon való tanulás lehetővé teszi az alapmodell számára egy új, ígéretes modell létrehozását a későbbi fázisokban, a ChestX-ray14 adatkészleten. Ebben a fázisban 198 628 MIMIC-CXR képet használtunk a tanításhoz és 49 657-et a validáláshoz. A képzést 50 iteráción keresztül végeztük, és a validálási AUC alapján választottuk ki a legjobb súlyokat tartalmazó állapotot.

4.3.3. Hiperoptimalizálás

A harmadik fázisban, ahogy az 4.1. ábrán látható, újra csak ResNet50V2-t használtuk, a 2. fázisban tanított súlyokkal, az architektúra teljes osztályozó részét eltávolítottuk és ezt használtuk kiindulópontként a Hyperband algoritmusunkban. A mélytanulási modellfejlesztésünk ezen fázisában a hiperparaméterek optimalizálására összpontosítottunk, hogy javítsuk a modell teljesítményét a konkrét feladatban. A hiperparaméter-optimalizálás kulcsfontosságú, mivel olyan modellbeállítások hangolását jelenti, amelyeket nem az adatokból tanulunk meg; ezek azonban jelentősen befolyásolhatják a tanulási folyamatot és a végső modell pontosságát.

Az optimalizáláshoz több kulcsfontosságú általános hiperparamétert azonosítottunk: veszteségfüggvény, optimalizáló beállítások és tanulási ráta. A veszteségfüggvény kiválasztása különösen kritikus annak meghatározásában, hogy a modell hogyan értelmezi a képzés során a megjósolt és a tényleges értékek közötti különbséget. Modellünkhöz kétféle veszteségfüggvényt vettünk figyelembe: BCE és SFL. Az SFL további paraméterekkel, nevezetesen α és γ jár, amelyek finomhangolást igényelnek. Ebben a fázisban is hasonlóan a 2.-hoz már csak és kizárólag a SGD optimalizáló algoritmust használjuk, viszont a hiperparaméter optimalizálás során optimalizáljuk az algoritmus *momentum* paraméterét. Az a paraméter arról ismert, hogy hatékonyan navigál a paramétertérben; a momentum kifejezés

segít felgyorsítani az optimalizálót a megfelelő irányba (ld. 1. algoritmus). A tanulási ráta az egyik legfontosabb hiperparaméter a neurális hálózatok képzésében. A cél a lépésméret meghatározása minden egyes iterációnál, miközben a veszteségfüggvény minimuma felé haladunk. Különböző tanulási rátákkal kísérleteztünk: 0, 1, 0, 01, 0, 001 és 0, 0001, hogy meghatározzuk a modellünk számára optimális rátát.

Ezek a paraméterek közvetlenül a tanulási folyamatra hatnak, általában ezeket szokták optimalizálni. Jelen fejezetben túllépve ezen, mi egy teljesen új osztályozó részt szeretnénk létrehozni a ResNet50V2 számára. Az egyetlen kötelező réteg az utolsó volt; egy sűrű réteget kellett használnunk 14 egységgel és egy szigmoid aktivációs függvénnyel. Ezen kívül teljesen szabad kezünk van, hogy a kinyert jellemzőréteg és ez a réteg közé milyen további rétegeket helyezünk. A következő táblázat összefoglalja a hiperparamétereket (4.2. táblázat) és a hozzájuk tartozó tartományokat vagy értékeket, amelyeket az optimalizálási folyamat során vizsgáltunk. Az első és a második oszlopban a hiperparaméterek optimalizálásában részt vevő paraméterek és tartományok szerepelnek. A Hyperopt1 , a Hyperopt2 és a Hyperopt3 a folyamat néhány ígéretes konfigurációját mutatta. Ezeket a 4. fázisban a teljes adatkészletre újratanítottuk. A Froze until paraméter azt jelzi, hogy a ResNet50V2 hány blokkja lett befagyasztva. A Dense 1,2 (a szám az egységeket jelenti, és a ReLU aktiválási függvényt használja) és a Dropout 1,2 (a szám a kiesési arányt jelenti) a modell osztályozási ré-

széhez hozzáadott extra rétegeket jelöli. A veszteség, az optimalizáló és a tanulási ráta saját tartományokkal rendelkezett. Ebben a fázisban ötszörös keresztvalidálást alkalmaztunk a ChestX-ray14 képeknek csak 0,2 képével, amelyet 13 843 képen képeztünk ki, és 3461 képen validáltunk. A Hyperbandnél 15 cikluson keresztül optimalizáltunk, és amennyiben 5 cikluson belül nem volt javulás, megállítottuk a tanulást.

4.2. táblázat. Hiperparaméter tartományok és kiválasztott értékek.

| Hiperparaméter megnevezése | Lehetséges értékek | $H_{yperopt}_1$ | $H_{yperopt}_2$ | $H_{yperopt}_3$ |
|---------------------------------|---------------------------------------|-----------------------------------------|-----------------|-----------------|
| Lezárt blokkok | block2, block3, block4, block5 | nincs | block 5 | block3 |
| Kiegészítő sűrű réteg 1 | nincs, 64, 128, 512 | 128 | 512 | 128 |
| Kiegészítő kiesési réteg 1 | 0,0-0,8 | 0,4 | 0,4 | 0,4 |
| Kiegészítő sűrű réteg 2 | nincs, 64, 128, 512 | 64 | 128 | nincs |
| Kiegészítő kiesési réteg 2 | 0,0-0,8 | 0,5 | 0 | |
| Veszteségfüggvény | BCE, SFL | SFL ($\alpha = 0, 25, \gamma = 2, 0$) | BCE | BCE |
| Optimalizáló | SFL (momentum: nincs, 0,5, 0,9, 0,99) | 0,9 | 0,9 | 0,5 |
| Tanulási ráta | 0,1, 0,01, 0,001, 0,0001 | 0,1 | 0,01 | 0,01 |
| Paraméterek száma | | 23 836 173 | 24 681 229 | 23 828 749 |
| Tanítható paraméterek száma | | 23 790 733 | 1 116 429 | 22 337 293 |
| Nem tanítható paraméterek száma | | 45 440 | 23 564 800 | 1 491 456 |

4.3.4. Optimalizált hálózat tesztelése

Az utolsó fázisban az optimalizált hálózatot a hivatalos ChestX-ray14 tanító adathalmazon tanítottuk. Ötszörös keresztvalidálást alkalmaztunk, 69 219 tanító és 17 305 validáló képpel. Az tanítást 30 cikluson keresztül futtattuk, és a validációs algoritmus segítségével választottuk ki a legjobb állapotot. A 4.3. táblázatban bemutatott összes modellkonfigurációt a hivatalos ChestX-ray14 tesztkészleten értékeltük, amely 25 596 képet tartalmaz. Néhány más, ebben a fázisban használt modellt is mellékelünk, hogy bemutassuk a javulás alakulását (4.3. táblázat).

4.3. táblázat. Eredmények különböző konfiguráción a ChestX-ray14 adatkészleten (átlagos AUC és szórás).

| | ImageNet Transzfer ¹ | MIMIC-CXR Transzfer ¹ | <i>Hyperopt</i> ₁ | <i>Hyperopt</i> ₂ | <i>Hyperopt</i> ₃ |
|-----------|------------------------------------|-------------------------------------|------------------------------|------------------------------|------------------------------|
| Atel. | 0,7632 ± 0,0115 | 0,8207 ± 0,0143 | 0,7604 ± 0,0076 | 0,8063 ± 0,0078 | 0,8287 ± 0,0023 |
| Card. | 0,6781 ± 0,0396 | 0,7162 ± 0,0190 | 0,8761 ± 0,0031 | 0,8798 ± 0,0061 | 0,8914 ± 0,0024 |
| Cons. | 0,7943 ± 0,0197 | 0,8200 ± 0,0161 | 0,7383 ± 0,0148 | 0,7920 ± 0,0011 | 0,8094 ± 0,0041 |
| Edema | 0,7829 ± 0,0220 | 0,8280 ± 0,0234 | 0,8340 ± 0,0072 | 0,8835 ± 0,0098 | 0,8995 ± 0,0044 |
| Effu. | 0,8725 ± 0,0037 | 0,9227 ± 0,0101 | 0,8229 ± 0,0053 | 0,8746 ± 0,0077 | 0,8931 ± 0,0002 |
| Emph. | 0,7239 ± 0,0092 | 0,7919 ± 0,0290 | 0,8658 ± 0,0085 | 0,8577 ± 0,0079 | 0,8847 ± 0,0033 |
| Fib. | 0,5977 ± 0,0183 | 0,7243 ± 0,0307 | 0,8039 ± 0,0080 | 0,7871 ± 0,0161 | 0,8086 ± 0,0027 |
| Hernia | 0,5774 ± 0,0409 | 0,6709 ± 0,0664 | 0,7571 ± 0,0991 | 0,8338 ± 0,0542 | 0,8130 ± 0,0378 |
| Infilt. | 0,6528 ± 0,0118 | 0,6968 ± 0,0079 | 0,6968 ± 0,0039 | 0,6803 ± 0,0162 | 0,7047 ± 0,0023 |
| Mass | 0,7215 ± 0,0130 | 0,7704 ± 0,0181 | 0,7836 ± 0,0113 | 0,8258 ± 0,0210 | 0,8518 ± 0,0020 |
| Nodule | 0,6564 ± 0,0161 | 0,7232 ± 0,0155 | 0,7207 ± 0,0095 | 0,7356 ± 0,0174 | 0,7588 ± 0,0025 |
| P. Thick. | 0,7129 ± 0,0336 | 0,7925 ± 0,0038 | 0,7585 ± 0,0121 | 0,7840 ± 0,0179 | 0,8121 ± 0,0025 |
| Pneum. | 0,7010 ± 0,0151 | 0,8027 ± 0,0227 | 0,7094 ± 0,0060 | 0,7156 ± 0,0138 | 0,7304 ± 0,0072 |
| PTX | 0,7478 ± 0,0143 | 0,8657 ± 0,0079 | 0,8512 ± 0,0073 | 0,8392 ± 0,0138 | 0,8646 ± 0,0012 |
| Átlag | 0,7130 ± 0,0020 | 0,7819 ± 0,0047 | 0,7842 ± 0,0007 | 0,8068 ± 0,0027 | 0,8250 ± 0,0023 |

¹ Eredeti ResNet50V2 egy darab dense réteggel.

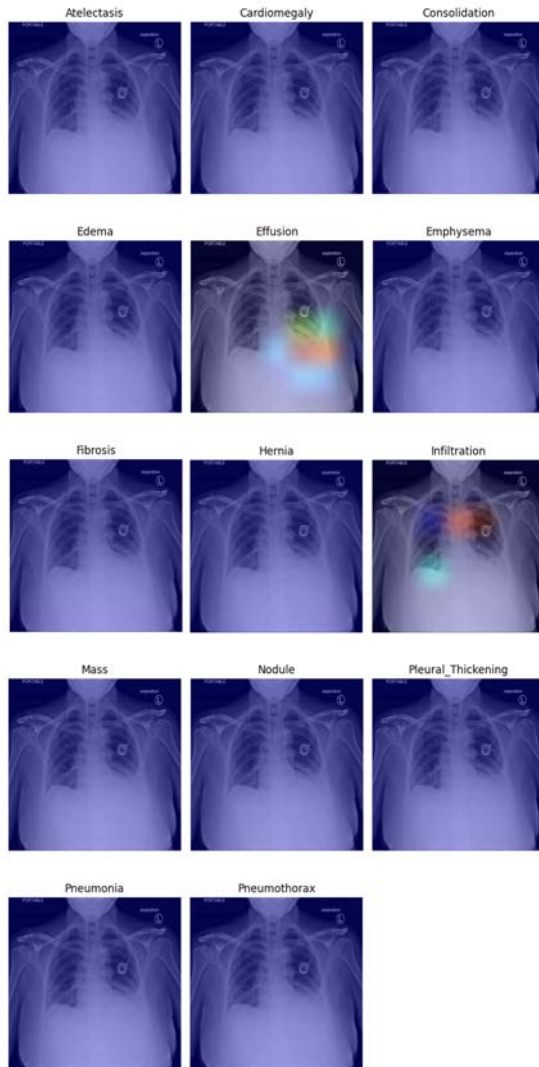
Az ImageNet transzfer oszlop a kindulási állapotot mutatja csak az utolsó réteget változtattuk meg, hogy illeszkedjen a ChestX-ray14 adatokhoz. A MIMIC-CXR transzfer oszlop a 2. fázisból származó, előre megtanult súlyokkal mutatott teljesítményt tartalmazza, ennek a két oszlopnak az egyéb konfigurációja megegyezik az 1. fázis legjobb konfigurációjával. A $Hyperopt_1$, a $Hyperopt_2$ és a $Hyperopt_3$ a hiperoptimalizálási folyamat során három jelentősen eltérő konfigurációjú teszt eredményeit tartalmazza.

4.3.5. Vizualizálás

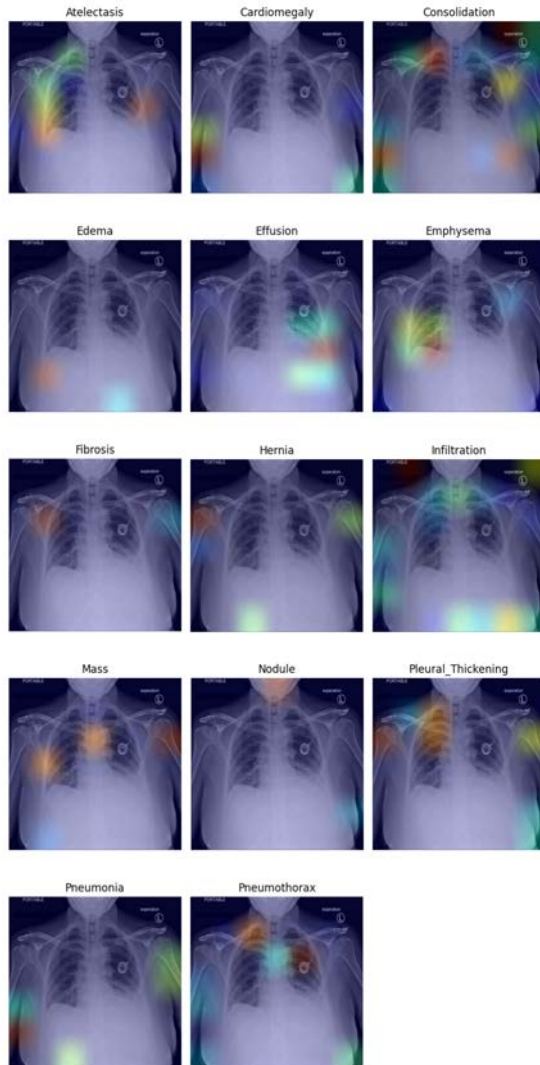
Ebben a szakaszban bemutatjuk a tanult jellemzők vizuális értelmezését a Grad-CAM [41] technika segítségével. A Grad-CAM olyan technikát kínál, amellyel a CNN-alapú modellek széles köréből származó döntések "vizuális magyarázatát" lehet előállítani, növelve azok átláthatóságát és értelmezhetőségét. A Grad-CAM bármely címke (pl. "Effúzió" a jelenlegi kontextusban) gradienseit felhasználja, amelyek az utolsó konvolúciós rétegbe áramlanak, hogy egy durva lokalizációs térképet hozzon létre. Ez a térkép kiemeli a képen a címke előrejelzése szempontjából kulcsfontosságú régiókat [41]. A korábbi módszerekkel ellentétben, a Grad-CAM különböző CNN-architektúrákra alkalmazható, anélkül, hogy újratanításra vagy architektúra-módosításra lenne szükség.

A 4.4. ábra az alapértelmezett osztályozófejet mutatja be. Az első sor első képén az átfedő bordaívek miatt tévesen atelektázist véltek felfedezni.

Mivel a bal oldali hemidiafragmát a megnagyobbodott szívárnyék eltakarta, az első sor ötödik képén a rekeszizom várt szintje alatt (késsel) van jelölve a folyadékgyülem. Az első sor hatodik képén eemfizéma volt feltételezhető, valószínűleg a tüdőalapok általános sűrűségének jelentős különbsége miatt. A második sor harmadik képén a megnagyobbodott aortacsomót tévesen térfoglalásnak vélték. A második sor hetedik képén a jobb apexben pneumothoraxot feltételeztek. Ezt a területet gyakran nehéz megítélni a kezdő radiológusoknak; mivel a bronchovascularis mintázat halvány, a pneumothorax téves gyanúját keltheti.



4.3. ábra. Grad-CAM térkép az új javasolt osztályozóval.



4.4. ábra. Grad-CAM térkép az alap osztályozóval.

Az 4.3. ábrán bemutatjuk a javasolt továbbfejlesztett modellünk vizuális ábrázolását, amelyben a mellhártyagyulladást az első sor ötödik képén jelöljük, amit egy radiológus megerősített. A második sor második képén azonban tévesen infiltrációt sugallnak, és helyette az aortacsomót jelölték (piros színnel). Az infiltráció kevésbé feltételezhető, de valószínű a következő esetekben a tüdő jobb alsó részén. A bal oldali folyadékgyülem mögött valószínűleg infiltráció is áll; ezt az algoritmus nem javasolta. A kardiomegáliát szintén nem javasolták, bár vizuálisan megerősíthető.

4.4. Összefoglalás

A bemutatottak szerint a javasolt módszerünk ígéretes AUC arányt mutat, amely számos konkurens modellt felülmúl ezen a kutatási területen. Figyelemre méltó, hogy csak két modell ért el jobb AUC arányt, mint a miénk (4.4. táblázat); azonban a módszereik jelentősen eltérnek a mi megközelítésünktől. Yan és mtsai. [42] határoló dobozokat használtak az érdeklődésre számot tartó régiókra (ROI) való összpontosításhoz, ami kétségtelenül segíti a tanulási folyamatot. A mi tanulmányunk azonban szándékosan az emberi beavatkozás minimalizálására összpontosít. Ez a megközelítés különösen fontos a nagy adathalmazok esetében, ahol a ROI-k kézi annotálása nem lehetséges az amúgy is túlterhelt radiológusok jelentős idő- és szakértelem igénye miatt. Baltruschat és mtsai. [43] további, nem képi jellemző-

ket használtak. Ők a beteg nemét és életkorát adták hozzá jellemzőként egy osztályozóhoz, amely jelentősen javította a nemtől vagy életkortól függő elváltozásokat. Továbbá egy jobb eredményt hozott a Kufel és mtsai. [44] által 2023. szeptember 22-én közzétett tanulmány, amely figyelemre méltóan magas AUC-t mutat; fontos megjegyezni, hogy eredményeik egyéni adatfelosztáson alapulnak. Ez az egyedi adatkezelés lehetetlenné teszi az összehasonlítást azok számára, akik a hivatalosat használják. Az adatfeldolgozási módszerek közötti különbség hangsúlyozza a kutatási protokollok szabványosításának fontosságát a tanulmányok közötti tisztességes és értelmes összehasonlítások biztosítása érdekében. Tanulmányunk kiváló alapot nyújt a terület jövőbeli kutatásaihoz. Az egyik ilyen lehetséges út a hardveres korlátozások kezelése és leküzdése. Ez megnyitja az utat a számításigényesebb sémákkal, például a ResNetV2 101 és 152 modellekkel való kísérletezéshez, vagy a vizsgálatokhoz magasabb hipertér létrehozásához. Ezenkívül megnyitja a kutatás egy másik szegmensét a különböző hálózati architektúrák erősségeit összeadó hiperoptimális együttes módszerek kifejlesztésére. Összefoglalva, kutatásunk hozzájárul az orvosi képelemzés jelenlegi megértéséhez és az orvosi képelemzésben használt módszerekhez, s megteremti a terepet az innovatív jövőbeli vizsgálatokhoz, amelyek tovább feszegethetik ezen a gyorsan fejlődő területen elérhető határokat.

4.4. táblázat. Eredményeink összehasonlítása különböző tanulmányokkal.

| | Wang és mtsai. [31] | Yao és mtsai. [47] | Shen és mtsai. [48] | Tang és mtsai. [32] | Guan és mtsai. [49] | Yan és mtsai. [42] ¹ | Baltruschat és mtsai. [43] ² | Kufel és mtsai. [44] ³ | Aktuális modell |
|--------------------|---------------------------|--------------------------|---------------------------|---------------------------|---------------------------|---------------------------------------|-----------------------------------------------|-----------------------------------------|--------------------|
| Atelectasis | 0,700 | 0,733 | 0,766 | 0,765 | 0,792 | 0,792 | 0,763 | 0,817 | 0,828 |
| Cardiomegaly | 0,810 | 0,856 | 0,801 | 0,887 | 0,879 | 0,881 | 0,875 | 0,911 | 0,891 |
| Consolidation | 0,703 | 0,711 | 0,787 | 0,728 | 0,758 | 0,760 | 0,749 | 0,815 | 0,809 |
| Edema | 0,805 | 0,806 | 0,820 | 0,848 | 0,850 | 0,848 | 0,846 | 0,908 | 0,899 |
| Effusion | 0,759 | 0,806 | 0,797 | 0,819 | 0,824 | 0,842 | 0,822 | 0,879 | 0,893 |
| Emphysema | 0,833 | 0,842 | 0,773 | 0,906 | 0,909 | 0,942 | 0,895 | 0,935 | 0,884 |
| Fibrosis | 0,786 | 0,743 | 0,765 | 0,818 | 0,832 | 0,833 | 0,816 | 0,824 | 0,808 |
| Hernia | 0,872 | 0,755 | 0,748 | 0,875 | 0,906 | 0,934 | 0,937 | 0,890 | 0,813 |
| Infiltration | 0,661 | 0,673 | 0,751 | 0,689 | 0,694 | 0,710 | 0,694 | 0,716 | 0,704 |
| Mass | 0,693 | 0,718 | 0,760 | 0,814 | 0,831 | 0,847 | 0,820 | 0,853 | 0,851 |
| Nodule | 0,669 | 0,777 | 0,741 | 0,755 | 0,766 | 0,811 | 0,747 | 0,771 | 0,758 |
| Pleural Thickening | 0,684 | 0,724 | 0,759 | 0,765 | 0,778 | 0,808 | 0,763 | 0,812 | 0,812 |
| Pneumonia | 0,658 | 0,684 | 0,778 | 0,729 | 0,726 | 0,740 | 0,714 | 0,769 | 0,730 |
| Pneumothorax | 0,799 | 0,805 | 0,800 | 0,850 | 0,858 | 0,876 | 0,840 | 0,898 | 0,864 |
| Átlag | 0,745 | 0,760 | 0,775 | 0,803 | 0,814 | 0,830 | 0,806 | 0,842 | 0,825 |

¹Kiegészítő fókuszálási régió adatokat használ.

²Egyéb nem kép alapú beteg információkat használ.

³Nem a hivatalos teszt készleten tesztelték.

5. fejezet

Ensemble módszerek alkalmazása többcímkes osztályozásra

Az előző fejezetekben bemutatott munkánk, beleértve a sajátunkat és az irodalmi áttekintésekben taglaltakat is egy CNN-modellt használtak a röntgenfelvételek címkézésére. Számos más valós probléma azonban azt mutatja, hogy egyetlen monolitikus rendszer kapacitása sem biztos, hogy elegendő azok megoldására. Ezt felismerve mind a természetes, mind a mesterséges rendszerek olyan megközelítéseket alkalmaznak, amelyek több, egymással összekapcsolt alrendszer együttműködésére támaszkodnak a komplexitás csökkentése és az összetett kihívások hatékony megoldása érdeké-

ben. A neurális hálózati ensemble (együttesek) létrehozásának módszertana változó. A leggyakrabban használt megközelítés egy kétlépcsős (offline) folyamat, amelyet Sharkey és mtsai. [50, 51] írták le először. Az első lépésben egyedi hálózatokat hoznak létre, majd bizonyos módon kombinálják őket. Ez a folyamat általában úgy történik, hogy minden egyes hálózatot külön-külön és a többitől függetlenül tanítják. Ennek az eljárásnak az egyik legnagyobb hátránya, hogy a képzés során elvész az egyes hálózatok közötti interakció lehetősége. Az együttes kialakításának ezen szakaszaiban nincs lehetőség a visszacsatolásra, ami azt jelenti, hogy az egyes hálózatok kialakítása és kombinálása között nincs információáramlás. Ez azt eredményezheti, hogy egyes egymástól függetlenül kialakított hálózatok kevésbé járulnak hozzá az együttes teljesítményéhez. Az egylépcsős (online) modellek olyan fejlett gépi tanulási megközelítések, amelyekben az egyes modellek tagjait egyidejűleg tanítják egyetlen hálózatban. Ezzel lehetővé téve, hogy a tagok közötti interakciók és visszajelzések közvetlenül befolyásolják egymást [52, 53, 54]. Ez a megközelítés segít kihasználni a modellek közötti szinergiákat, amivel a tagok a tanulási folyamat során tanulnak egymástól, ami növeli az előrejelzési pontosságot és a modell általános robusztusságát. Számos cikk szemlélteti, hogy az együttes megközelítés használható osztályozási problémákban [55, 56, 57]. Megannyi technika létezik arra, hogy a modelleket együttes modellé kapcsoljuk össze, az egyszerű számtani átlagoktól, a súlyozott átlagoktól, a fejlettebb többségi szavazáson

át a fejlett másodlagos tanuláson keresztül [58, 59, 60, 61, 62, 63]. Másodlagos tanulásnak nevezzük, amikor a tagokat egy neurális hálózattal kombináljuk össze, ennek a kombináló hálózatnak a bemenete a tagok kimeneteinek konkatenációja, tetszőleges számú belső neurális réteggel rendelkezik, kimenete pedig a célfüggvényünk kimenete. Néhány más cikk, kifejezetten az együttes CNN technikák orvosi felhasználását mutatja be [64, 65, 66]. A diverzitás biztosítása terén kritikusan fontosságú az egylépéses együttes modellek optimalizálása, ami jelentősen javíthatja a modell általánosítási képességét. Egyes technikák mind az egylépéses, mind a többlépéses együttes modellekhez is alkalmazhatók, beleértve a különböző előfeldolgozási és adatbővítési technikákat [52] és a különböző architektúrák használatát [67]. Az egyik legpraktikusabb és leghatékonyabb megközelítés, a diverzitás elérésének a sokféleségre vonatkozó kényszerítés közvetlen integrálása a veszteségfüggvénybe. A szakirodalomban ez a megközelítés először olyan munkában jelent meg, amely a veszteségfüggvénybe korrelációs büntetőfaktorokat vezetett be [68]. Céljük az volt, hogy a gradiens redukciós algoritmust diverzebb együttesek előállítására ösztönözzön. Opitz és mtsai. [53] egy speciálisan gépi tanuláshoz kifejlesztett veszteségfüggvényt mutattak be, amelynek célja egy változatos és alacsony korrelációjú együttes modell létrehozása volt. A koszinusz hasonlóság büntető terminusként való használata szintén hatékony módszernek bizonyult a modell diverzitásának növelésére [52]. Korábbi kutatásainkban bemutattuk a Pearson-féle

korreláció használatát kategorikus képosztályozási feladatokhoz az együttes tanulás sokszínűségének növelése érdekében [69]. Ezek a megközelítések lehetővé teszik, hogy jobban kihasználjuk az együttes modellekben rejlő lehetőségeket, növelve a tanulási folyamat hatékonyságát és a végső modell pontosságát. Amenyiben kilépünk az online ensemblek világából találunk olyan megoldást amely kombinált veszteség függvényt használ. A Han mtsai.[Han] által tárgyalt veszteségfüggvény bemutat egy kombinált veszteséget amely Euklidészi távolság alapú veszteségeket kombinál, mégpedig olyan módon hogy a tag függvények veszteségei és a hálózat teljes veszteségének arányát súlyozza, Boosting technikát alkalmazva.

A többcímkes osztályozási feladatok sajátossága, hogy előfordulhatnak olyan esetek, amikor egy adott mintához nem társul címke, ami nullvektort eredményez. Egy másik speciális eset, amikor a vektorban több egyes (beleértve az összes egyeseket is) van, ami több (vagy az összes) címke jelenlétére utal. Ez a helyzet különösen nagy kihívást jelent, amikor büntető technikákat alkalmazunk a modell diverzitásának növelése érdekében. A Pearson-féle korreláció vagy a koszinusz hasonlóság, amelyek általában hatékony eszközök a modellek közötti kapcsolatok, s azok sokféleségének mérésére nem bizonyulnak praktikusnak, ha a címkék teljes hiánya is szóba jöhet. Ebben az esetben ezek a metrikák torzított vagy irreleváns eredményeket produkálhatnak, mivel nem képesek kezelni a nullvektorok által jelzett speciális eseteket.

Az egyes modellek veszteségeinek figyelembevétele lehetővé teszi, hogy a végső döntési mechanizmus, ne csak az együttes általános teljesítményére összpontosítson, hanem az egyes tagok hozzájárulásának értékelésére is. Ezáltal, ha az együttes jelentősen eltérő architektúrákból áll, a modell diverzitása növekszik, amivel az együttes nem csak "átlagolt" döntést hoz, hanem figyelembe veszi az egyes tagok egyedi, potenciálisan eltérő jeleit is, ami javítja a modell robusztusságát és alkalmazkodóképességét a különböző címkézési helyzetekhez.

Ebben a fejezetben egy új online ensemble technikát mutatunk be képosztályozási feladatokhoz. Újításunk kulcsa egy speciálisan kifejlesztett kombinált veszteségfüggvény, amely egyedileg kiszámítja a teljes ensemble modell veszteségét, és kiszámítja a résztvevő modellkomponensek veszteségét. Ezt az ötletet úgy valósítjuk meg, hogy több modellt egyesítünk, és variáljuk az összevonásukat, hogy egy új hálózati architektúra alkalmazásával növeljük az osztályozási pontosságot. Ehhez több előre betanított CNN-ből létrehozunk egy ensemble-t, és különböző kombinációs módszereket alkalmazunk. Az egyik módszer a különálló CNN-ek összekapcsolása egy további, az eredeti CNN-ek kimeneti rétege után beillesztett, teljesen összekapcsolt rétegen keresztül. Emellett egy másik stratégiát is alkalmazunk, ahol a modellkomponensek kimeneti rétegeit címké-súlyozott átlagolással kombináljuk, ahol az átlagszámítás paraméterei megtanulhatók. Végül olyan módszert alkalmazunk, amelyben bár minden egyes tagmodell-

nek saját kimeneti rétege van, de ezeket kizárólag az egyesített veszteségfüggvény kiszámítására használjuk. Ezzel szemben az összekapcsoláshoz egyesítjük a tagkomponensek jellemzőkinyerő rétegeit, és az új, teljesen összekapcsolt réteget helyezzük rájuk. Ez a többrétegű megközelítés lehetővé teszi, hogy kihasználjuk az egyes modellkomponensek erősségeit, miközben javítjuk az osztályozási teljesítményt és növeljük a modell általánosítási képességét.

5.1. Módszertan

5.1.1. Architektúrák

Az elmúlt néhány évben a konvolúciós neurális hálózatok váltak az iparág vezető technológiájává a digitális képfeldolgozásban mint például mintafelismerés, szegmentálása és osztályozás. Ezek a hálózatok képesek megtanulni a konvolúciós szűrők paramétereit, és olyan értelmes és magasabb szintű jellemzőket nyernek ki, amelyek a képek különböző kategóriái közötti megkülönböztetésre használhatók. A képzési fázis nagy számítási teljesítményt és több ezer kézzel címkézett képet igényel. A javasolt hálózatok elég általánosak ahhoz, hogy alacsony hibahatárral alkalmazhatók legyenek különböző osztályozási feladatokra. Ez az igény vezetett a közelmúltban több új CNN-architektúra kifejlesztéséhez, mint például az InceptionV3 [71], a ResNet [36], a DenseNet [37], a MobileNet [72], a

RegNet [73] és az Xception [74]. Ezek előre betanított modellek formájában is elérhetőek, eredetileg az ImageNet adathalmazon betanítva. Eme modellek transzfer tanulás keretében használhatók, ahol a súlyok és torzítások segítségével finomhangolhatjuk a modelleket a saját adatainkra. Ha elegendő annotált kép áll rendelkezésre egy nulláról induló modell létrehozásához, akkor a hálózati paramétereket véletlenszerűen inicializálhatjuk. Mivel a ResNet és a DenseNet az előző fejezetben ismertetésre került, itt nem részletezzük.

RegNet

A RegNet [73], a Facebook AI által bevezetett hálózati struktúrák sorozata a tervezés során a szimulálhatóságot és a skálázhatóságot helyezi előtérbe, és olyan modelleket mutat be, amelyek mind a pontosság, mind a hatékonyság szempontjából finomhangoltak. Ezek között a modellek között van a RegNetX016, amely az igényekhez igazodik. A RegNetX sorozat részeként a RegNetX016 a csoportosított konvolúciók felhasználásával a hatékonyságra összpontosít, hogy összhangot találjon a követelmények és a teljesítmények között. Egyes architektúrákkal ellentétben nem tartalmaz Squeeze and Excitation (SE) blokkokat, így ez egy egyszerű, könnyen skálázható és telepíthető kialakítás, különösen korlátozott képességű hardvereken.

MobileNet

A MobileNetV2 [72], a MobileNet sorozat kiegészítése, amely a hatékonyságot előtérbe helyezve a látási feladatok mobil eszközökön való kiemelkedő teljesítményét szolgálja. Ez a módszer invertált reziduumokat és lineárisan szűk keresztmetszeteket használ a hálózaton belüli információáramlás fokozására, miközben csökkenti a terhelést. A mélyreható konvolúciók felhasználásával ezen kialakítás a modell méretének és összetettségének racionalizálására törekszik, így ideális a számítási képességekkel rendelkező mobil és beágyazott eszközökön történő telepítéshez. A MobileNetV2 jellegéből adódóan a különböző képfelismerési feladatok során is megőrzi a pontosságot, s hatékonyan egyensúlyozva a teljesítmény és a hatékonyság között. Alkalmazkodóképessége és hatékonysága miatt népszerű választássá vált olyan valós idejű alkalmazásokhoz, mint a tárgyak felismerése és az arcfelismerés erőforrásokkal rendelkező környezetben.

Xception

Az Xception [74] egy új megközelítés bevezetésével átalakítja a konvolúciós neurális hálózatokat. Ez olyan konvolúciókat használ, amelyek szétválasztják a csatornák közötti és a térbeli korrelációk kezelését. Ez az architektúra 36 rétegből áll, amelyek 14 szakaszba vannak szervezve és a hatékonyságra és a teljesítményre összpontosítanak, s a kapcsolatok fel-

használásával a gradiens eltűnési problémák kezelésére alkalmasak. Az Xceptiont a kutatásunk egyik elemeként integráltuk, ezáltal hatékony lesz a jellemzők kinyerésére és az erőforrások optimalizálására. Az Xception beépítése döntő szerepet játszott abban, hogy modellünk képes legyen a képek pontos osztályozására. Bizonyított sokoldalúsága a különböző feladatokban igazolja döntésünket, hogy értékeljük a képalapú osztályozási törekvéseink pontosságára és hatékonyságára gyakorolt hatását.

InceptionV3

A Szegedy és mtsai. [71] által kifejlesztett InceptionV3 architektúra számos új réteget és blokkstruktúrát vezet be, amelyek optimalizálják a hálózat teljesítményét és hatékonyságát. Az egyik legfontosabb újítás az úgynevezett asszimmetrikus konvolúció alkalmazása, amelyek különböző méretű szűrőket használnak párhuzamosan egy adott rétegen belül. Ez lehetővé teszi a hálózat számára, hogy különböző szintű és méretű jellemzőket ragadjon meg, így pontosabb osztályozást érhet el, kevesebb számítási teljesítmény alkalmazása mellett.

5.1.2. Ensemble módszerek

Ha egy ensemble architektúrát szeretnénk létrehozni, akkor választhatunk az egylépéses és a kétlépéses (online és offline) ensemble módszerek

között. Az egy lépéses módszernél az összes tagot ugyanabban a folyamatban kell tanítanunk, a kétlépésesnél viszont külön-külön is taníthatjuk a tagokat, és utána fuzionálhatjuk őket. Ebben a tanulmányban csak az egy lépéses technikákkal foglalkozunk. A mély ensemble-hálózat kiemeli a mélytanulási technikák használatát, különösen az integrációs módszerhez tartozó képzési rétegek használatát az ensemble döntéshozatali folyamatának javítása érdekében. Jelen esetben ensemble modelleknek két fő típusát különböztetjük meg. Mindkét típus M különböző tagból állhat. Az első típus az, amikor minden tagnak saját kimeneti rétege van és a modellek egyedi előrejelzéseit egy (vagy több) teljesen összekapcsolt réteg segítségével összekapcsolják, ami ezeket az előrejelzéseket kombinálja. Ezt a típust nevezzük osztályszintűnek (ld. a 5.1. a) ábrát). A második típus az, ahol csak a gerinchálózatok által kivont jellemzők vannak a konkatenációhoz felhasználva, és ezek a tagok nem rendelkeznek saját kimeneti réteggel. Az összekapcsolt jellemzők felett egy teljesen összekapcsolt réteget alkalmazunk, hasonlóan az osztályszintű együtteshez. Ez a megközelítés lehetővé teszi a kivont jellemzők közvetlen felhasználását az előrejelzések integrálására és finomítására. Ezt a típust nevezzük jellemzőszintű ensemble-nek (ld. a 5.1. b ábrát). Mindkét típusú együttes célja, hogy kihasználja a különböző modellek által nyújtott információk sokféleségét, ezáltal javítva az előrejelzések pontosságát és a modell általánosítási képességét. Az osztályszintű együttes a modellspecifikus döntések közötti sokféleséget használja

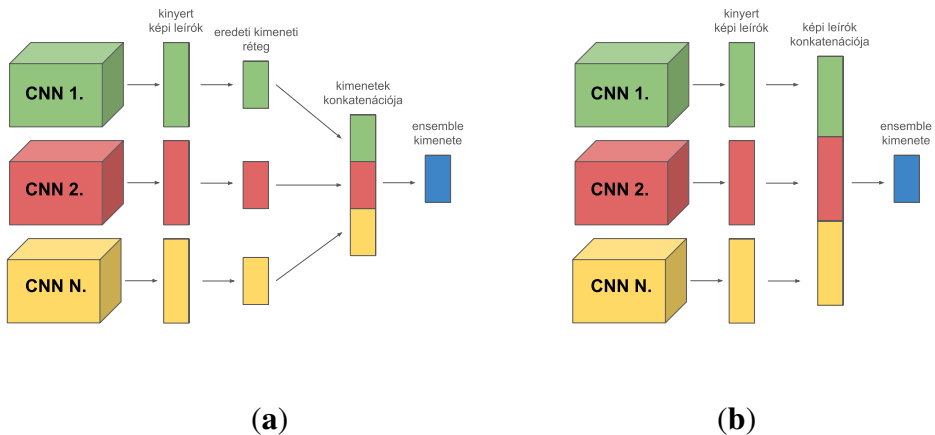
ki. Ezzel szemben a jellemzőszintű ensemble a kivont jellemzők sokféleségére összpontosít, így robusztusabb döntési mechanizmust biztosít a végső osztályozásban. A teljesség kedvéért a tagok kimeneti rétegeinek súlyozott átlagát is referenciaként használtuk. Ebben az esetben nem használunk teljesen összekapcsolt réteget a konkatenációhoz. Egy olyan egyéni réteget használtunk, amely a kimeneti címkék átlagát számítja ki, tanítható súlyokkal:

$$\hat{y} = \sum_{k=1}^M w_k \cdot y_k \quad (5.1)$$

A (5.1)-ban \hat{y} a súlyozott átlagolt előrejelzést jelképezi, M pedig az ensemble modellben lévő tagok teljes számát. Az y_k a k -adik modell előrejelzési kimenetére utal, míg w_k a k -adik taghoz rendelt súly. A többcímkes esetben az y_k és w_k a címkék valószínűségeit tartalmazó vektorokat jelent. A súlyok használata lehetővé teszi, hogy dinamikusan kiigazítsuk az egyes modellek befolyását a végső előrejelzésre. Elismerve, hogy az egyes modellek megbízhatóbban vagy relevánsabban járulhatnak hozzá bizonyos előrejelzésekhez, mint mások. Ez a rugalmasság a súlyozott átlagú megközelítés döntő előnye az egyszerű átlaggal szemben, hogy lehetővé teszi az egyes modellek erősségeinek célzottabb kihasználását. A súlyok optimális halmazának (w_k) meghatározása kritikus szempont, a súlyozott átlag stratégiájának hatékony alkalmazása szempontjából. Esetünkben a súlyok taníthatóak, és ezeket a súlyokat az adatokból automatikusan tanítottuk;

természetesen a lehetőség adott ezen súlyvektorok manuális megadására is.

A súlyozott átlag réteg és a dense réteg közötti különbség tisztázása érdekében fontos megjegyezni, hogy a dense réteg csak akkor felelne meg egy súlyozott átlagnak, ha a súlymátrixában csak a főátlóban található 0-tól különböző elem. Azonban a stochasztikus tanítási folyamat miatt erre kevés esély van. Ezért az általunk felvázolt súlyozott átlag réteg mintegy kényszeríti a modellt erre a konfigurációra, biztosítva, hogy az aggregálás során minden egyes bemeneti jellemző megfelelő súlyt kapjon.



5.1. ábra. Különböző típusú együttések: (a) osztályszintű; (b) jellemzőszintű.

5.1.3. Javasolt új veszteségfüggvény

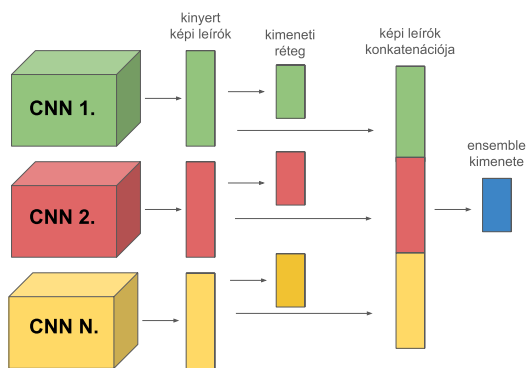
A többcímkes osztályozási feladatok jellemzője, hogy lehetnek olyan minták, amelyeknek egyáltalán nincs címkéjük, és így nullvektort kapunk. Ezért a kategorikus osztályozásban vizsgált, általánosan használt büntető kifejezéses megközelítés nem ad jó eredményeket. Ezért más oldalról közelítettük meg a problémát, az egyes osztályozók részvételét súlyozva. A modell sokszínűségének növelése érdekében hasznos az egyes modellek veszteségeit a közös veszteségszámításba integrálni. Ez a megközelítés nemcsak az együttes összteljesítményét veszi figyelembe, hanem az egyes komponensek egyéni hozzájárulását is, ami növeli a modell alkalmazkodóképességét és robusztusságát különböző címkézési helyzetekben. Ez a fajta diverzitás különösen fontos, amikor különböző architektúrájú modellekből álló együtteseket használunk, amivel lehetővé tesszük, hogy az együttes többféle nézőpontból közelítse meg az adott problémát, ezáltal javítva az osztályozási pontosságot. Ily módon a modellünk hatékonyabban integrálja az egyes komponensek által szolgáltatott információkat, optimalizálva az osztályozási teljesítményt. Különösen a több címkével ellátott képosztályozási feladatokban várunk jelentős pontosság javulást a megoldásunktól. Így megközelítésünk nemcsak a tanulási folyamat hatékonyságát javítja, hanem jelentősen növeli a modellünk általánosítási képességét is, biztosítva, hogy jobban alkalmazkodjon a különböző kihívásokhoz. Ezen tanul-

mányban bemutatott mély együttes modell tréningezhető fejrétegekkel rendelkezik, és egy magas komplexitású, sok paraméterrel rendelkező struktúra, ami különösen nagy kihívást jelent a hatékony tanulás szempontjából. Az egylépeses együttes módszer esetében ezeket a paramétereket egyszerre optimalizálhatjuk egy összehangolt tanulási lépésben, amelynek célja az együttes modell teljesítményének javítása. Módszerünk a modell egyes tagjainak veszteségeit kombinálja az együttes kimeneti rétegen számított veszteséggel. Ez a stratégia lehetővé teszi számunkra, hogy finomhangoljuk az egyes modellkomponensek hozzájárulását az együttes döntéshozatali folyamatához, ezáltal növelve a tanulás hatékonyságát és javítva az osztályozási pontosságot. Tehát a megközelítésünk nemcsak új dimenziókat nyit meg a mélytanulási modelltanulásban, hanem javítja a többcímkes képosztályozási feladatokban elérhető eredményeket is, kihasználva az együttes modellekben rejlő potenciált. Az általunk javasolt kompozit veszteségfüggvényt a következőképpen számítjuk ki:

$$\mathcal{L}_{total} = L_{Multi-BCE}(y, \hat{y}_{out}) + \lambda \cdot \frac{1}{M} \sum_{k=1}^M L_{Multi-BCE}(y, \hat{y}_k), \quad (5.2)$$

ahol M a taghálózatok számát jelöli, $L_{Multi-BCE}$ a többcímkes bináris keresztentropia veszteségfüggvényt (2.8), y a kép valódi címkéjét jelenti, \hat{y}_{out} az együttes kimenetét jelöli, y_k pedig a k -adik tag kimenetét, és végül λ a tagok veszteségét szabályozza. λ súlyozás egy eszközt biztosít a külön-

böző modellkomponensek egyéni hozzájárulásának kezelésére az ensemble modellre gyakorolt hatását illetően.



5.2. ábra. Módosított jellemzőszintű ensemble az új javasolt veszteségfüggvényhez.

Az osztályszintű esetben a tagok veszteségeinek kiszámításához a tagok teljesítményét használtuk. A jellemzőszintű együttes esetében azonban módosítani kellett a modellt, és teljesen összekapcsolt rétegeket kellett hozzáadnunk kimeneti réteggé a veszteség kiszámításához (ld. a 5.2 ábrát). Ebben az esetben a teljesen összekapcsolt rétegek a hagyományos használatán túl kritikus szerepet játszanak mivel kifejezetten csak a veszteség kiszámításához használjuk őket, amit az (5.2) egyenlet részletezi. Ez a beállítás lehetővé teszi a tagmodellek erősségeinek mélyebb integrálását, optimalizálva az ensemble modell általános előrejelző képességét a speciális veszteségszámítás kihasználásával a tanulás hatékonyságának és a modell

pontosságának növelése érdekében.

5.2. Eredmények

5.2.1. A különböző CNN tagok kiértékelése

Átfogó áttekintést készítettünk a DenseNet121, InceptionV3, MobileNetV2, RegNetX016, ResNet50V2 és Xception teljesítményéről. Minden kiértékelést a Chest X-ray14 adathalmaz 25%-án végeztünk, a címkék eredeti eloszlását a lehető legnagyobb mértékben megőrizve. Ötszörös keresztvalidálást alkalmaztunk, az AUC értékek átlagát a 5.1. táblázatban mutatjuk be. Ezeket a méréseket 32 elemű kötegekkel, 30 cikluson keresztül és két optimalizáló konfigurációval végeztük. A mérések során azt a megközelítést alkalmaztuk, hogy a hivatalos teszt adatkészleten kizárólag a legutolsó lépésben teszteltük az optimális modellünket így azon semmilyen extra optimalizációt nem végeztünk, az előtte lévő szakaszokban csak a tanító adathalmazt használtuk minden esetben ötszörös keresztvalidációt alkalmazva.

Az első optimalizáló konfiguráció az Adam (ld. 2. algoritmus) 0,001-es tanulási rátával és $\beta_1 = 0,9$ és $\beta_2 = 0,999$, a második egy SGD (ld. 1. algoritmus) *momentum* = 0 és 0,01-es tanulási rátával. Láttuk, hogy minden modell stabil teljesítményt mutat (ld. tab:finding-members. táblázatban) az SGD-nél, ezért a két létrehozott együttes az SGD-t használja op-

timalizáló tanulási rátájaként 0,01-nél és 0 momentummal. Megfigyeltük, hogy az SGD konzisztensebb teljesítményt mutat, mint az Adam optimalizáló, ezért a továbbiakban csak az SGD-t használjuk.

| | Tanítás | | Validálás | |
|-------------|--------------------|--------------------|--------------------|--------------------|
| | Adam | SGD | Adam | SGD |
| DenseNet121 | 0,7796 ± 0,0120 | 0,8082 ± 0,0045 | 0,7569 ± 0,0260 | 0,7693 ± 0,0084 |
| ResNet50V2 | 0,7885 ± 0,0044 | 0,8146 ± 0,0033 | 0,7428 ± 0,0304 | 0,7631 ± 0,0066 |
| MobileNetV2 | 0,8244 ± 0,0068 | 0,7882 ± 0,0054 | 0,7356 ± 0,0245 | 0,7590 ± 0,0057 |
| InceptionV3 | 0,8162 ± 0,0069 | 0,7685 ± 0,0030 | 0,7546 ± 0,0144 | 0,7436 ± 0,0058 |
| RegNetX016 | 0,8259 ± 0,0008 | 0,7277 ± 0,0024 | 0,7669 ± 0,0162 | 0,7274 ± 0,0055 |
| Xception | 0,9186 ± 0,0030 | 0,6987 ± 0,0033 | 0,6998 ± 0,0143 | 0,6969 ± 0,0023 |

5.1. táblázat. Tagok önálló teljesítménye (AUC)..

5.2.2. Az ensemble modellek vizsgálata

Az átfogó értékelés után két különböző ensemble architektúrát javasoltunk, betartva egy kritikus korlátozást: az ensemble modell teljes memóriafogyasztása nem haladhatja meg a 14 GB-ot, a memóriahasználatot 32 köteg mérettel és 224×224 képmérettel számoltuk. Az 5.2. táblázat összehasonlító áttekintést nyújt ezekről a javasolt ensemble modellekről, bemutatva a paraméterek teljes számát és a memóriaigényt az egyes konfigurá-

ciók esetében. Az első javasolt ensemble architektúrában az egyes modellek a rétegek közötti speciális kapcsolatokra támaszkodnak (DenseNet121), vagy egyszerre különböző skálákon dolgoznak (InceptionV3), illetve olyan modellek, amelyek segítségével mélyebb hálózatokat építhetünk anélkül, hogy problémákba ütköznénk (ResNet50V2). A DenseNet121 annyiban különleges, hogy minden egyes réteg kapcsolódik az előzőekhez, így jobban kihasználja az adatokat, és kevesebb paraméterrel is futtatható. Az InceptionV3 több méretben dolgozik, így könnyebben felismeri a különböző méretű dolgokat. A ResNet50V2 pedig segít mélyebbé tenni a hálózatokat anélkül, hogy a tanulás során elveszítenénk az irányítást.

5.2. táblázat. Memóriahasználát és tanítható paraméterek száma a javasolt együttesek esetén.

| | Modell | Tanítható paraméterek száma | Memóriahasználás | Összes tanítható paraméterek száma | Teljes memória használat |
|-----------------------------|-------------|-----------------------------|------------------|------------------------------------|--------------------------|
| <i>Ensemble₁</i> | ResNet50V2 | 25 568 360 | 3855 MB | | |
| | DenseNet121 | 7 978 856 | 6102 MB | 57 364 568 | 13 954 MB |
| | InceptionV3 | 23 817 352 | 3997 MB | | |
| <i>Ensemble₂</i> | Xception | 22 855 952 | 7710 MB | | |
| | RegNetX016 | 9 190 136 | 3120 MB | 35 550 960 | 13 496 MB |
| | MobileNetV2 | 3 504 872 | 2666 MB | | |

A második javasolt ensemble modellben a hatékonyságra és a szisztematikus tervezésre összpontosítottunk (RegNetX016), és kis számítási teljesítménnyel dolgoztunk (MobileNetV2), valamint új konvolúciós módsze-

reket használtunk (Xception). A RegNetX016 egy olyan hálózat, amelyet úgy terveztek, hogy hatékonyan működjön, és jól működik a különböző feladatokkal. A MobileNetV2 szintén speciális megoldásokat alkalmaz, hogy jól működjön mobil eszközökön, míg az Xception újítást jelent a térbeli és a csatorna-információk külön kezelésével, ami javítja a teljesítményt. Ez a két ensemble modell tehát különböző architektúrákat használ a több címkés osztályozási problémák megoldására, és úgy véljük, hogy sokféleségük javítja az ensemble modellek teljesítményét. Az 5.2. táblázat szerinti alapvizsgálatként az $Ensemble_1$ és az $Ensemble_2$ ensemble rendszereket valósítottuk meg. Mindkét javasolt modellben a tagok összekapcsolása jellemző szintű és osztály szintű, illetve egyszerű súlyozott átlagolási módszertan mentén történik (5.1).

Természetesen itt is ötszörös keresztvalidálást alkalmaztunk. Az eredmény az 5.3. táblázatban látható. Ezeket a konfigurációkat a ChestX-ray14 adathalmazon végeztük el, többcímkes bináris keresztentropia használatával. Következő lépésként megvizsgáljuk, hogy az általunk javasolt veszteségfüggvény hogyan javíthatja ezen ensemble modellek teljesítményét (ld. a 5.1.3 szakaszt), különösen a λ paraméter szerepének vizsgálatával. Ez a λ paraméter befolyásolja az egyes modellek egyedi veszteségeinek súlyát a végső veszteségszámításban. A λ értékének növelésével azt akaruk tesztelni, hogy az egyes modellek által okozott veszteségek mennyire dominálnak a teljes veszteségben, és ez hogyan befolyásolja az együttes

| | | Tanítási AUC | Validálási AUC |
|-----------------------------|-----------------|--------------------|--------------------|
| <i>Ensemble₁</i> | Jellemzőszintű | 0.8036 ± 0.0039 | 0.7662 ± 0.0109 |
| | Súlyozott átlag | 0,7783 ± 0,0053 | 0,7542 ± 0,0078 |
| | Osztályszintű | 0,6931 ± 0,0106 | 0,7029 ± 0,0061 |
| <i>Ensemble₂</i> | Jellemzőszintű | 0,7675 ± 0,0028 | 0,7549 ± 0,0041 |
| | Súlyozott átlag | 0,7275 ± 0,0047 | 0,7210 ± 0,0047 |
| | Osztályszintű | 0,6703 ± 0,0051 | 0,6748 ± 0,0069 |

5.3. táblázat. *Ensemble₁* és *Ensemble₂* eredmények az új javasolt veszteség függvény nélkül.

modell teljesítményét. Eredményeink az 5.4. és a 5.5. táblázatban azt mutatják, hogy amikor a λ értékét mérsékelten növeltük 0,5-ről 1,0-ra, majd 2,0-ra, a modell teljesítménye jelentősen javult a kombinált veszteség nélküli alapváltozathoz képest. Ez arra utal, hogy egy megfelelően beállított λ érték, amely nagyobb hangsúlyt fektet az egyes modellekben az egyéni veszteségekre, hozzájárulhat a modell általánosíthatóságának javításához. Ez lehetővé teszi, hogy a modell jobban kihasználja az egyes komponensek által nyújtott egyedi jellemzőket és információkat, ami elősegíti a pontosabb osztályozást. Izgalmas jelenség, hogy az osztályszintű együttesek nem érnek el jobb eredményeket az általunk javasolt veszteségek használata esetén. Jelenleg figyelmünk kizárólag a jobb eredmények elérésére irányul. Amikor azonban a λ értéke irreálisan magasra, 5,0-ra vagy 10,0-ra nőtt, a modell teljesítményének csökkenését figyeltük meg a hagyományos modellhez képest. Ez a jelenség arra utal, hogy túl magas λ -értékek esetén a modell túlságosan az egyes modellek egyéni veszteségeire összpontosít, ami korlátozhatja a tanulási folyamatot és negatívan befolyásolhatja a modell alkalmazkodóképességét. Ezek az eredmények rávilágítanak arra, hogy a λ paraméter gondos beállítása elengedhetetlen a modell optimális teljesítményének eléréséhez. A λ paraméter finomhangolása lehetővé teszi, hogy pontosan szabályozzuk az egyes modellek hozzájárulását a végső veszteségszámításhoz, így segítve egy kiegyensúlyozott és hatékonyabb osztályozási rendszer kialakítását. Ez a megközelítés új perspektívákat nyit

a többcímkes ensemble modellek továbbfejlesztése előtt, hangsúlyozva az egyes modellkomponensek veszteségeinek szerepét a végső teljesítmény javításában.

| | λ értéke | Tanítási AUC | Validálási AUC |
|-----------------|------------------|---------------------------------------|---------------------------------------|
| Jellemzőszintű | nincs | 0,8036 \pm 0,0039 | 0,7662 \pm 0,0109 |
| | 0,5 | 0,8391 \pm 0,0038 | 0,7700 \pm 0,0040 |
| | 1,0 | 0,8680 \pm 0,0020 | 0,7718 \pm 0,0065 |
| | 2,0 | 0,9076 \pm 0,0037 | 0,7407 \pm 0,0094 |
| | 5,0 | 0,9397 \pm 0,0042 | 0,7608 \pm 0,0092 |
| | 10,0 | 0,9490 \pm 0,0029 | 0,7534 \pm 0,0088 |
| Osztályszintű | nincs | 0,6931 \pm 0,0106 | 0,7029 \pm 0,0061 |
| | 0,5 | 0,7257 \pm 0,0011 | 0,6871 \pm 0,0038 |
| | 1,0 | 0,7457 \pm 0,0059 | 0,6877 \pm 0,0029 |
| | 2,0 | 0,7745 \pm 0,0011 | 0,6832 \pm 0,0063 |
| | 5,0 | 0,8101 \pm 0,0018 | 0,6806 \pm 0,0069 |
| | 10,0 | 0,8245 \pm 0,0019 | 0,6760 \pm 0,0025 |
| Súlyozott átlag | nincs | 0,7783 \pm 0,0053 | 0,7542 \pm 0,0078 |
| | 0,5 | 0,8454 \pm 0,0042 | 0,7775 \pm 0,0017 |
| | 1,0 | 0,8725 \pm 0,0158 | 0,7821 \pm 0,0061 |
| | 2,0 | 0,9159 \pm 0,0020 | 0,7847 \pm 0,0041 |
| | 5,0 | 0,9474 \pm 0,0017 | 0,7809 \pm 0,0085 |
| | 10,0 | 0,9516 \pm 0,0024 | 0,7565 \pm 0,0058 |

5.4. táblázat. $Ensemble_1$ eredményei az új veszteségfüggvénnyel különböző λ értékek mellett .

| | λ értéke | Tanítási AUC | Validálási AUC |
|-----------------|------------------|---------------------------------------|---------------------------------------|
| Jellemzőszintű | nincs | 0,7675 \pm 0,0028 | 0,7549 \pm 0,0041 |
| | 0,5 | 0,7965 \pm 0,0054 | 0,7735 \pm 0,0055 |
| | 1,0 | 0,8157 \pm 0,0018 | 0,7717 \pm 0,0127 |
| | 2,0 | 0,8447 \pm 0,0022 | 0,7766 \pm 0,0114 |
| | 5,0 | 0,8806 \pm 0,0018 | 0,7690 \pm 0,0054 |
| | 10,0 | 0,9100 \pm 0,0015 | 0,7641 \pm 0,0207 |
| Osztályszintű | nincs | 0,6703 \pm 0,0051 | 0,6748 \pm 0,0069 |
| | 0,5 | 0,7016 \pm 0,0031 | 0,6853 \pm 0,0099 |
| | 1,0 | 0,7145 \pm 0,0047 | 0,6898 \pm 0,0095 |
| | 2,0 | 0,7339 \pm 0,0042 | 0,6928 \pm 0,0069 |
| | 5,0 | 0,7677 \pm 0,0052 | 0,6935 \pm 0,0104 |
| | 10,0 | 0,7977 \pm 0,0065 | 0,6943 \pm 0,0070 |
| Súlyozott átlag | nincs | 0,7275 \pm 0,0047 | 0,7210 \pm 0,0047 |
| | 0,5 | 0,7952 \pm 0,0013 | 0,7694 \pm 0,0170 |
| | 1,0 | 0,8217 \pm 0,0022 | 0,7825 \pm 0,0029 |
| | 2,0 | 0,8610 \pm 0,0027 | 0,7924 \pm 0,0077 |
| | 5,0 | 0,9071 \pm 0,0036 | 0,7838 \pm 0,0093 |
| | 10,0 | 0,9422 \pm 0,0012 | 0,7810 \pm 0,0073 |

5.5. táblázat. $Ensemble_2$ eredményei az új veszteségfüggvénnyel különböző λ értékek mellett .

5.2.3. Kombinálás transzfer tanulással

A korábbi 4. fejezetben gerinchálózati architektúrák képzésére összpontosítottunk a MIMIC-CXR adatkészlet felhasználásával a teljesen összekapcsolt rétegek építésére. A jelenlegi tanulmány továbbfejleszti ezt a megközelítést azáltal, hogy minden egyes tagot önállóan tanítottunk a MIMIC-

CXR adatkészleten, a korábban meghatározott kijelölt optimalizálók és ensemble-konfigurációk alkalmazásával. Ehhez a képzési fázishoz 256 elemű köteggel és összesen 50 ciklust alkalmaztunk, egy 250 000 frontális mellkasi röntgenfelvételt tartalmazó adathalmazt használva az előre betanított modellek kifejlesztéséhez. SGD optimalizálót használtunk 0,01-es tanulási rátával és 0,9-es momentummal, kiválasztottuk a legjobb AUC-t a validációs adathalmazon. Ezután ezeket a hálózatokat kombináltuk össze a módosított veszteségfüggvényünkkel, ahol λ 2,0 volt, és súlyozott átlagos fejet választottuk. A hivatalos ChestX-ray14 teszt adatkészletet használunk, hogy összehasonlítható eredményeket kapjunk más kutatók munkáival. Az eredmények a 5.6. táblázatban láthatók, a jobb eredményt félkövérrrel jelöltük.

5.6. táblázat. Eredmények előtanított tagokkal az új és a hagyományos veszteségfüggvényt használva .

| | <i>Ensemble</i> ₁ Nincs | <i>Ensemble</i> ₁ $\lambda = 2.0$ | <i>Ensemble</i> ₂ Nincs | <i>Ensemble</i> ₂ $\lambda = 2.0$ |
|--------------------|---------------------------------------|-------------------------------------------------|---------------------------------------|-------------------------------------------------|
| Atelectasis | 0,8068 | 0,8323 | 0,8061 | 0,8258 |
| Cardiomegaly | 0,8856 | 0,9014 | 0,8916 | 0,9009 |
| Consolidation | 0,7915 | 0,8123 | 0,7850 | 0,8128 |
| Edema | 0,8951 | 0,9040 | 0,9056 | 0,9053 |
| Effusion | 0,8777 | 0,8915 | 0,8763 | 0,8883 |
| Emphysema | 0,8435 | 0,8918 | 0,8566 | 0,8785 |
| Fibrosis | 0,7561 | 0,8076 | 0,7770 | 0,7924 |
| Hernia | 0,7037 | 0,8628 | 0,5685 | 0,7138 |
| Infiltration | 0,6487 | 0,6489 | 0,6780 | 0,6562 |
| Mass | 0,8164 | 0,8676 | 0,7975 | 0,8511 |
| Nodule | 0,7421 | 0,7773 | 0,7258 | 0,7602 |
| Pleural Thickening | 0,7756 | 0,8208 | 0,7759 | 0,8076 |
| Pneumonia | 0,6997 | 0,7465 | 0,7209 | 0,7300 |
| Pneumothorax | 0,8220 | 0,8651 | 0,8346 | 0,8400 |
| Átlag | 0,7904 | 0,8307 | 0,7857 | 0,8116 |

5.2.4. Összehasonlítás a korszerű technikákkal

A kutatásunk során elvégzett kiértékelés és tesztek eredményei alapján (ld. 5.7. táblázat) az első ensemble modellünk (*Ensemble*₁) felülmúlta a legkorszerűbb modelleket. Ezzel szemben a második ensemble modellünk (*Ensemble*₂), bár nem érte el az első modell kiemelkedő teljesítményét, számos más rendszernél jobban teljesített. Ezek az eredmények jelentős mérföldkövet jelentenek a többcímű osztályozási feladatokhoz való megközelítésünkben, különösen a mélytanulás és az ensemble modellek területe-

tén. Az *Ensemble*₁ sikere rávilágít arra, hogy innovatív módszertanunk - különösen a különböző modellek kombinálásával és egy speciálisan kifejlesztett veszteségfüggvény használatával - új utakat nyit a képosztályozási feladatok hatékonyságának javítására. Ez arra ösztönöz bennünket, hogy tovább vizsgáljuk az ensemble modellekben rejlő lehetőségeket, különösen a többcímkes kontextusban. Az *Ensemble*₂ eredményei, bár nem érik el az első csoport kiemelkedő teljesítményét, megerősítik azt a nézetet, hogy az ensemble modellek használata jelentős javulást jelenthet számos jelenlegi modellhez képest. Ez azt sugallja, hogy még a szuboptimális konfigurációk is versenyképes eredményeket érhetnek el, különösen, ha figyelembe vesszük a többcímkes problémák összetettségét és kihívásait.

5.7. táblázat. Eredménynek összehasonlítása a legkorszerűbb eredményekkel. .

| | Tang | Guan | Yan | Balruschat | Kufel | 4. fejezet | |
|--------------------|-------------------|-------------------|--------------------------------|--------------------------------|--------------------------------|-----------------------|-----------------------|
| | és mtsai. [32] | és mtsai. [34] | és mtsai. [42] ¹ | és mtsai. [43] ² | és mtsai. [44] ³ | Ensemble ₁ | Ensemble ₂ |
| | | | | | | $\lambda = 2.0$ | $\lambda = 2.0$ |
| Atelectasis | 0,765 | 0,792 | 0,792 | 0,763 | 0,817 | 0,828 | 0,826 |
| Cardiomegaly | 0,887 | 0,879 | 0,881 | 0,875 | 0,911 | 0,891 | 0,900 |
| Consolidation | 0,728 | 0,758 | 0,760 | 0,749 | 0,815 | 0,809 | 0,812 |
| Edema | 0,848 | 0,850 | 0,848 | 0,846 | 0,908 | 0,899 | 0,905 |
| Effusion | 0,819 | 0,824 | 0,842 | 0,822 | 0,879 | 0,893 | 0,888 |
| Emphysema | 0,906 | 0,909 | 0,942 | 0,895 | 0,935 | 0,884 | 0,878 |
| Fibrosis | 0,818 | 0,832 | 0,833 | 0,816 | 0,824 | 0,808 | 0,792 |
| Hernia | 0,875 | 0,906 | 0,934 | 0,937 | 0,890 | 0,813 | 0,713 |
| Infiltration | 0,689 | 0,694 | 0,710 | 0,694 | 0,716 | 0,704 | 0,656 |
| Mass | 0,814 | 0,831 | 0,847 | 0,820 | 0,853 | 0,851 | 0,851 |
| Nodule | 0,755 | 0,766 | 0,811 | 0,747 | 0,771 | 0,758 | 0,760 |
| Pleural Thickening | 0,765 | 0,778 | 0,808 | 0,763 | 0,812 | 0,812 | 0,807 |
| Pneumonia | 0,729 | 0,726 | 0,740 | 0,714 | 0,769 | 0,730 | 0,730 |
| Pneumothorax | 0,850 | 0,858 | 0,876 | 0,840 | 0,898 | 0,864 | 0,840 |
| Mean | 0,803 | 0,814 | 0,830 | 0,806 | 0,842 | 0,825 | 0,811 |

¹Kiegészítő fókuszálási régió adatokat használ.

²Egyéb nem kép alapú beteg információkat használ.

³Nem a hivatalos teszt készleten tesztelték.

5.3. Összefoglalás

Az ensemble modelljeink eredményeit látva megállapíthatjuk, hogy jelentős javulást lehet elérni még a többcímkes problémák esetében is. Ezért további vizsgálatokat javasolunk, beleértve a modellek szélesebb körének vizsgálatát, valamint még több modell konfigurálását 4-6-8-ig. Javasoljuk továbbá, hogy mélyebben ássunk bele a lehetséges módszerekbe, beleértve az új veszteségfüggvényeket és azok hatását a modellek teljesítményére. A modell kölcsönhatása még szélesebb körű szemlélete és a kombinált megközelítésből származó sokféleség révén még nagyobb teljesítményt feltételezünk a többcímkes képosztályozásban. Ezen túlmenően a további modellek integrálása gazdagabb területet kínál a közös dinamikájuk feltárásához és a közös teljesítményük optimalizálásához. Kutatásunk rávilágít a többcímkes együttes modellek jelentős ígéretére a képosztályozási kihívások kezelésében, és új utakat mutat be a jövőbeli kutatások számára. Ezek az eredmények megerősítik elkötelezettségünket, hogy több időt és erőforrást szenteljünk ennek a többcímkes területnek, mivel óriási potenciál rejlik benne.

6. fejezet

Összefoglalás

Ebben a disszertációban a többcímkes osztályozás problémakörét vizsgáltuk meg mellkasröntgen felvételeken. A mellkasröntgen felvételek elemzése kritikus jelentőségű a különböző tüdő, szív és sok más betegség diagnosztizálásában. A többcímkes osztályozás lehetővé teszi, hogy egy adott röntgenfelvételen több betegség jelenlétét is detektálja. Ez különösen fontos, mivel egyetlen felvételen egyszerre több betegség is előfordulhat. Az általunk vizsgált megközelítések a mélytanulási modellek hatékonyságára és pontosságára építenek, különös tekintettel a CNN-ek alkalmazásával.

A 3. fejezetben egy többcímkes osztályozót építünk az alapoktól. Ennek az osztályozónak az alapját konvolúciós rétegek képezik. Különböző mélységű hálózatokat vizsgálunk képesek-e ezt az összetett kép osztályozási problémát abszolválni. A javasolt modellek képesek a nulláról megta-

nulni az alapvető címkéket de sajnos a state-of-the-art megoldásokkal nem vették fel a versenyt, így a további fejezetben ismert architektúrákkal haladtunk tovább. A fejezetben tárgyalt eredményeket a [75] publikációban mutattuk be.

A 4. fejezetben, egyelemű osztályozót alkottunk transzfer tanulás és hiperparaméter optimalizálás segítségével. A transzfer tanulás lehetővé teszi, hogy egy előzetesen betanított modellt, amely nagy mennyiségű adatot használt, újra felhasználjunk egy másik, hasonló feladatra. Itt ismert CNN gerinceket: a ResNet, VGG és DenseNet architektúrákat alkalmaztunk alapmodellként. Megvizsgáltuk, képesek-e röntgenfelvételek több-címkés osztályozására, majd transzfer tanulás segítségével inicializáltuk a legígéretesebbet. Erre az előtanított hálózatra építettünk egy teljesen új több rétegből álló osztályozó részt. Ez a módszer összehasonlításként szolgált ChestX-ray14 adatkészleten és eredményesebbnek bizonyult mint más csoportok által publikált eredmények. A [76] publikációban mutattuk be a fejezet eredményeit.

A 5. fejezetben továbblépve az egyelemű hálózatokon, egy többelemű ensemble osztályozót fejlesztettünk ki, amely különböző architektúrájú CNN osztályozók kombinációjából áll. Az ensemble modell célja, hogy javítsa az egyes modellek pontosságát és robusztusságát azáltal, hogy azok döntéseit kombinálja.

Az általunk javasolt új kombinált veszteségfüggvény az egyes tagmo-

dellek veszteségeit kombinálja a hálózat teljes veszteségével. Az így létrejött új ensemble osztályozó nem csak a mások által javasolt módszereket, hanem az általunk használt eddigi módszereket is felülmúlta. A harmadik rész eredményeit a [77] publikáció tárgyalja, amely jelen dolgozat írásakor elbírálás alatt volt.

Az itt bemutatott megközelítések új lehetőségeket nyitnak a gépi tanulás és az orvosi képfeldolgozás területén. Az eredmények azt mutatják, hogy a fejlett mélytanulási technikák hatékonyan alkalmazhatók a komplex orvosi feladatok megoldására, és hozzájárulhatnak a későbbiekben a pontosabb és megbízhatóbb diagnózisokhoz, javítva ezzel a betegellátás minőségét.

Köszönetnyilvánítás

Köszönöm Istennek és a Debreceni Egyetemnek, hogy lehetőséget kaptam a tanulásra.

Köszönöm Dr. Harangi Balázs témavezető tanár úrnak, hogy végig támogatta a kutatásomat, és értékes tanácsaival, útmutatásával segített a disszertáció megírásában. Hálás vagyok a türelméért, szakértelméért és az iránymutatásáért, amelyek nélkül ez a munka nem jöhetett volna létre.

Köszönöm Prof. Dr. Pethő Attila, Prof. Dr. Baran Sándor, Prof. Dr. Sztrik János doktori iskola vezető professzor uraknak és Prof. Dr. Hajdu András és Prof. Dr. Kruppa András programvezető professzor uraknak, hogy lehetővé tették számomra, hogy az Informatikai Tudományok Doktori Iskolában tanulhattam.

Köszönöm volt témavezetőimnek Dr. Várterész Magda tanárnőnek, Prof. Dr. Vereb György professzor úrnak és Dr. Antal Bálint tanár úrnak, hogy egyetemi éveim alatt egyengették az utamat.

Köszönöm volt tanárimnak, külön köszöntettel Dr. Máté József tanár úrnak és Moczok Gyula tanár úrnak.

Külön köszönet illeti társszerzőimet, Dr. Tóth Gábort, Dr. Petró Mátyást, Dr. Rebenku Istvánt és Dr. Besenczi Renátót, akik hozzájárultak kutatásomhoz és segítettek annak megvalósításában.

Köszönöm édesanyámnak, Katonáné Katona Erzsébetnek, hogy megszeretette velem a matematikát, édesapámnak Katona Bélának, hogy folytonosan ösztönzött a jobb tanulmányi teljesítményre és végig támogattak

a tanulmányaim során. Köszönetet szeretnék mondani a családomnak és barátaimnak is, akik folyamatosan támogattak és bátorítottak a kutatásom során.

Külön köszönet illeti a feleségemet, Katonáné Bodnár Zsófiát, aki végtelen türelemmel és szeretettel támogatott az egész disszertációm írása alatt. Köszönöm, hogy mellettem álltál azokon a végtelen éjszakákon, amikor a kutatásaim rabjává váltam, és hogy elviselted a fáradtság okozta pillanatok, amikor már csak a te bátorításod adott erőt a folytatáshoz. Végül, de nem utolsósorban, szeretnék köszönetet mondani születendő gyermekemnek, aki már most inspirál és motivál arra, hogy mindennap a legjobbat nyújtsam.

Irodalomjegyzék

- [1] Rosenkrantz, A.B.; Hughes, D.R.; Duszak, R., Jr. The U.S. Radiologist Workforce: An Analysis of Temporal and Geographic Variation by Using Large National Datasets. *Radiology* **2016**, *279*, 175-184, doi:10.1148/radiol.2015150921.
- [2] Ruuskanen, O.; Lahti, E.; Jennings, L.C.; Murdoch, D.R. Viral pneumonia. *Lancet* **2011**, *377*, 1264–1275. [https://doi.org/10.1016/S0140-6736\(10\)61459-6](https://doi.org/10.1016/S0140-6736(10)61459-6).
- [3] Ali, F.; Harrington, S.; Kennedy, S.; Hussain, S. Diagnostic radiology in Liberia: A country report. *J. Glob. Radiol.* **2015**, *1*, 1–7. <https://doi.org/10.7191/jgr.2015.1020>.
- [4] Fatihoglu, E.; Aydin, S.; Gokharman, F.D.; Ece, B.; Kosar, P.N. X-ray use in chest imaging in emergency department on the basis of cost and effectiveness. *Acad. Radiol.* **2016**, *23*, 1239–1245. <https://doi.org/10.1016/j.acra.2016.05.008>.
- [5] Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July, 2017; pp. 3462–3471.
- [6] Kruger, R.P.; Townes, J.R.; Hall, D.L.; Dwyer, S.J.; Lodwick, G.S. Automated Radiographic Diagnosis via Feature Extraction and Classification of Cardiac Size and Shape Descriptors. *IEEE Trans. Biomed. Eng.* **1972**, *BME-19*, 174–186. <https://doi.org/10.1109/TBME.1972.324115>.
- [7] Lodwick, G.S.; Keats, T.E.; Dorst, J.P. The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer. *Radiology* **1963**, *81*, 185–200. <https://doi.org/10.1148/81.2.185>.
- [8] Meyers, P.H.; Nice, C.M.; Becker, H.C.; Nettleton, W.J.; Sweeney, J.W.; Meckstroth, G.R. Automated Computer Analysis of Radiographic Images. *Radiology* **1964**, *83*, 1029–1034. <https://doi.org/10.1148/83.6.1029>.
- [9] de Bruijne, M. Machine learning approaches in medical image analysis: From detection to diagnosis. *Med. Image Anal.* **2016**, *476*, 94–97. <https://doi.org/10.1016/j.media.2016.06.032>.

- [10] Johnson, A.E.W.; Pollard, T.J.; Berkowitz, S.J.; Greenbaum, N.R.; Lungren, M.P.; Deng, C.Y.; Mark, R.G.; Horng, S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci. Data* **2019**, *6*, 317. <https://doi.org/10.1038/s41597-019-0322-0>.
- [11] Johnson, A.E.W.; Pollard, T.J.; Greenbaum, N.R.; Lungren, M.P.; Ying Deng, C.; Peng, Y.; Lu, Z.; Mark, R.G.; Berkowitz, S.J.; Horng, S. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv* **2019**, arXiv:1901.07042.
- [12] Irvin, J.; Rajpurkar, P.; Ko, M.; Yu, Y.; Ciurea-Ilcus, S.; Chute, C.; Marklund, H.; Haghgoo, B.; Ball, R.; Shpanskaya, K.; et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In Proceedings of the AAAI Conference on Artificial Intelligence 2019, Honolulu, HI, USA, 27 January–1 February 2019; pp. 590–597.
- [13] Zhu, C.S.; Pinsky, P.F.; Kramer, B.S.; Prorok, P.C.; Purdue, M.P.; Berg, C.D.; Gohagan, J.K. The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial and Its Associated Research Resource. *J. Natl. Cancer Inst.* **2013**, *105*, 1684–1693. <https://doi.org/10.1093/jnci/djt281>.
- [14] Bustos, A.; Pertusa, A.; Salinas, J.M.; de la Iglesia-Vayá, M. PadChest: A large chest X-ray image dataset with multi-label annotated reports. *Med. Image Anal.* **2020**, *66*, 101797. <https://doi.org/10.1016/j.media.2020.101797>.
- [15] Kermany, D.; Zhang, K.; Goldbaum, M. Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data* **2018**, *2*, 651. <https://doi.org/10.17632/rscbjbr9sj>.
- [16] RSNA Challenge. <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pneumonia-detection-challenge-2018> (10 December 2023).
- [17] Vayá, M.d.I.I.; Saborit, J.M.; Montell, J.A.; Pertusa, A.; Bustos, A.; Cazorla, M.; Galant, J.; Barber, X.; Orozco-Beltrán, D.; García-García, F.; et al. BIMCV COVID-19+: A large annotated dataset of RX and CT images from COVID-19 patients. *arXiv* **2020**, arXiv:2006.01174.
- [18] Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, Physio-

Toolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, E215–E220, <https://doi.org/10.1161/01.cir.101.23.e215>.

- [19] Ruder, S: An overview of gradient descent optimization algorithms *arXiv* **2016**. <https://arxiv.org/abs/1609.04747>
- [20] Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. *MIT press* **2016**.
- [21] Diederik, P. K.; Jimmy B.: Adam: A Method for Stochastic Optimization *arXiv* **2017**. <https://arxiv.org/abs/1412.6980>
- [22] Peng, Y., Wang, X., Lu, L., Bagheri, M., Summers, R., Lu, Z. NegBio: a high-performance tool for negation and uncertainty detection in radiology reports. AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science, San Francisco, CA, USA, March 27–30 2017; pp. 188–196.
- [23] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... Ng, A. Y. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv* **2017**, arXiv:1711.05225.
- [24] Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**. <https://arxiv.org/abs/1409.1556>
- [25] Krizhevsky, A., Hinton, G. E. (n.d.). ImageNet Classification with Deep Convolutional Neural Networks. Advances in neural information processing systems, Lake Tahoe, NV, USA, December 3–8, 2012. vol 25.
- [26] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2015**, arXiv:1409.4842.
- [27] tubercnn Lakhani, P., Sundaram, B. (2017). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* **2017**, *284*:2, <https://doi.org/10.1148/radiol.2017162326>.
- [28] Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Comput.* **1997**; *9* (8): pp. 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [29] Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., Lyman, K. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv* **2017**, arXiv:1710.10501.

- [30] Irvin, J.; Rajpurkar, P.; Ko, M.; Yu, Y.; Ciurea-Ilcus, S.; Chute, C.; Marklund, H.; Haghighi, B.; Ball, R.; Shpanskaya, K.; et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In Proceedings of the AAAI Conference on Artificial Intelligence 2019, Honolulu, HI, USA, 27 January–1 February 2019; pp. 590–597.
- [31] Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Summers, R.M. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-Rays. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, UT, USA, 18 June–22 June 2018; pp. 9049–9058.
- [32] Tang, Y.; Wang, X.; Harrison, A.P.; Lu, L.; Xiao, J.; Summers, R.M. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In Proceedings of the International Workshop on Machine Learning in Medical Imaging (MLMI 2018), Granada, Spain, 16 September 2018; Springer: Berlin/Heidelberg, Germany 2018; Volume 1; pp. 249–258.
- [33] Yao, L.; Prosky, J.; Poblenz, E.; Covington, B.; Lyman, K. Weakly supervised medical diagnosis and localization from multiple resolutions. *arXiv* **2018**, arXiv:1803.07703.
- [34] Guan, Q.; Huang, Y. Multi-label chest X-ray image classification via category-wise residual attention learning. *Pattern Recognit. Lett.* **2020**, *130*, 259–266. <https://doi.org/10.1016/j.patrec.2018.10.027>.
- [35] Wang, H.; Wang, S.; Qin, Z.; Zhang, Y.; Li, R.; Xia, Y. Triple attention learning for classification of 14 thoracic diseases using chest radiography. *Med. Image Anal.* **2021**, *67*, 101846. <https://doi.org/10.1016/j.media.2020.101846>.
- [36] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**. <https://arxiv.org/abs/1512.03385>
- [37] Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993.
- [38] O’Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; et al. KerasTuner. <https://github.com/keras-team/keras-tuner> (17 November 2023).

- [39] Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *J. Mach. Learn. Res.* **2018**, *18*, 1–52.
- [40] Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 Augst 2019; pp 2623–2631. <https://doi.org/10.1145/3292500.3330701>.
- [41] Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359, <https://doi.org/10.1007/s11263-019-01228-7>.
- [42] Yan, C.; Yao, J.; Li, R.; Xu, Z.; Huang, J. Weakly Supervised Deep Learning for Thoracic Disease Classification and Localization on Chest X-rays. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Washington DC, USA, 29 August–1 September 2018; pp. 103–110. <https://doi.org/10.1145/3233547.3233573>.
- [43] Baltruschat, I.M.; Nickisch, H.; Grass, M.; Knopp, T.; Saalbach, A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. *Sci. Rep.* **2019**, *9*, 6381. <https://doi.org/10.1038/s41598-019-42294-8>.
- [44] Kufel, J.; Bielówka, M.; Rojek, M.; Mitęga, A.; Lewandowski, P.; Cebula, M.; Krawczyk, D.; Bielówka, M.; Kondol, D.; Bargiel-Łączek, K.; et al. Multi-Label Classification of Chest X-ray Abnormalities Using Transfer Learning Techniques. *J. Pers. Med.* **2023**, *13*, 1426. <https://doi.org/10.3390/jpm13101426>.
- [45] Chen, X.; Yao, L.; Zhou, T.; Dong, J.; Zhang, Y. Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. *Pattern Recognit.* **2021**, *113*, 107826. <https://doi.org/10.1016/j.patcog.2021.107826>.
- [46] Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; Mcalpine, D.; Zhang, Y. A Survey on Deep Learning-based Non-Invasive Brain Signals: Recent Advances and New Frontiers. *arXiv* **2020**, arXiv:1905.04149.
- [47] Yao, L.; Poblens, E.; Dagunts, D.; Covington, B.; Bernard, D.; Lyman, K. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv* **2018**, arXiv:1710.10501.

- [48] Shen, Y.; Gao, M. Dynamic Routing on Deep Neural Network for Thoracic Disease Classification and Sensitive Area Localization, 2018. *arXiv* **2018**, arXiv:1808.05744.
- [49] Guendel, S.; Grbic, S.; Georgescu, B.; Liu, S.; Maier, A.; Comaniciu, D. Learning to recognize abnormalities in chest X-rays with 552 location-aware dense networks. In Proceedings of the Iberoamerican Congress on Pattern Recognition (CIARP 2018), Madrid, Spain, 19–22 November 2018; Springer: Berlin/Heidelberg, Germany 2019; pp. 757–765.
- [50] Sharkey, Amanda J. C. On combining artificial neural nets. **Connection science** **1996**: 299-314.
- [51] Sharkey, A.J.C., Sharkey, N.E., Gerecke, U., Chandroth, G.O. (2000). The “Test and Select” Approach to Ensemble Combination. In: Multiple Classifier Systems. Multiple Classifier Systems 2000, Calgiari, Italy, 21-23 June 2000; Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg. https://doi.org/doi:10.1007/3-540-45014-9_3.
- [52] Dvornik, N.; Mairal, J.; Schmid, C. Diversity With Cooperation: Ensemble Methods for Few-Shot Classification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3722–3730. <http://dx.doi.org/10.1109/ICCV.2019.00382>
- [53] Opitz, M.; Waltner, G.; Possegger, H.; Bischof, H. Deep Metric Learning with BIER: Boosting Independent Embeddings Robustly. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, 42, 276–290. <http://dx.doi.org/10.1109/TPAMI.2018.2848925>
- [54] Zhang, L.; Shi, Z.; Cheng, M.M.; Liu, Y.; Bian, J.W.; Zhou, J.T.; Zheng, G.; Zeng, Z. Nonlinear Regression via Deep Negative Correlation Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, 43, 982–998. <http://dx.doi.org/10.1109/ACCESS.2019.2933670>
- [55] Zhang, B.; Qi, S.; Monkam, P.; Li, C.; Yang, F.; Yao, Y.D.; Qian, W. Ensemble Learners of Multiple Deep CNNs for Pulmonary Nodules Classification Using CT Images. *IEEE Access* **2019**, 7, 110358–110371. <http://dx.doi.org/10.1109/ACCESS.2019.2933670>

- [56] Kuehlkamp, A.; Pinto, A.; Rocha, A.; Bowyer, K.W.; Czajka, A. Ensemble of Multi-View Learning Classifiers for Cross-Domain Iris Presentation Attack Detection. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1419–1431. <http://dx.doi.org/10.1109/TIFS.2018.2878542>
- [57] Maarouf, A.A.; Hachouf, F. Transfer Learning-based Ensemble Deep Learning for Road Cracks Detection. In Proceedings of the 2022 International Conference on Advanced Aspects of Software Engineering (ICAASE), Constantine, Algeria, 17–18 September 2022; pp. 1–6. <http://dx.doi.org/10.1109/ICAASE56196.2022.9931581>
- [58] Khan, I.A.; Sajeeb, A.; Fattah, S.A. An Automatic Ocular Disease Detection Scheme from Enhanced Fundus Images Based on Ensembling Deep CNN Networks. In Proceedings of the 11th International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, 17–19 December 2020; pp. 491–494. <http://dx.doi.org/10.1109/ICECE51571.2020.9393050>
- [59] Li, W.; Liu, H.; Wang, Y.; Li, Z.; Jia, Y.; Gui, G. Deep Learning-Based Classification Methods for Remote Sensing Images in Urban Built-Up Areas. *IEEE Access* **2019**, *7*, 36274–36284. <http://dx.doi.org/10.1109/ACCESS.2019.2903127>
- [60] Chen, Y.; Wang, Y.; Gu, Y.; He, X.; Ghamisi, P.; Jia, X. Deep Learning Ensemble for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1882–1897. <http://dx.doi.org/10.1109/JSTARS.2019.2915259>
- [61] Minetto, R.; Pamplona Segundo, M.; Sarkar, S. Hydra: An Ensemble of Convolutional Neural Networks for Geospatial Land Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6530–6541. <http://dx.doi.org/10.1109/TGRS.2019.2906883>
- [62] Dong, S.; Feng, W.; Quan, Y.; Dauphin, G.; Gao, L.; Xing, M. Deep Ensemble CNN Method Based on Sample Expansion for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. <http://dx.doi.org/10.1109/TGRS.2022.3183189>
- [63] Alosaimi, N.; Alhichri, H. Fusion of CNN ensemble for Remote Sensing Scene Classification. In Proceedings of the 2020 3rd International Conference on Computer Applications Information Security (ICCAIS), Riyadh, Saudi Arabia, 19–21 March 2020; pp. 1–6. <http://dx.doi.org/10.1109/ICCAIS48893.2020.9096721>

- [64] Elmannai, H.; Saleh, H.; Algarni, A.D.; Mashal, I.; Kwak, K.S.; El-Sappagh, S.; Mostafa, S. Diagnosis Myocardial Infarction Based on Stacking Ensemble of Convolutional Neural Network. *Electronics* **2022**, *11*, 3976. <https://doi.org/10.3390/electronics11233976>
- [65] Mansoor, R.; Shah, M.A.; Khattak, H.A.; Mussadiq, S.; Rauf, H.T.; Ameer, Z. Detection of Diseases in Pandemic: A Predictive Approach Using Stack Ensembling on Multi-Modal Imaging Data. *Electronics* **2022**, *11*, 3974. <https://doi.org/10.3390/electronics11233974>
- [66] Mahanty, C.; Kumar, R.; Asteris, P.G.; Gandomi, A.H. COVID-19 Patient Detection Based on Fusion of Transfer Learning and Fuzzy Ensemble Models Using CXR Images. *Appl. Sci.* **2021**, *11*, 11423. <https://doi.org/10.3390/app112311423>
- [67] Harangi, B. Skin lesion classification with ensembles of deep convolutional neural networks. *J. Biomed. Inform.* **2018**, *86*, 25–32. <http://dx.doi.org/10.1016/j.jbi.2018.08.006>
- [68] Liu, Y.; Yao, X. Ensemble learning via negative correlation. *Neural Netw.* **1999**, *12*, 1399–1404. [http://dx.doi.org/10.1016/S0893-6080\(99\)00073-8](http://dx.doi.org/10.1016/S0893-6080(99)00073-8)
- [69] Harangi, B.; Baran, A.; Beregi-Kovacs, M.; Hajdu, A. Composing Diverse Ensembles of Convolutional Neural Networks by Penalization. *Mathematics*, **2023**, *11*, 4730. <https://doi.org/10.3390/math11234730>
- [Han] Han, S.; Meng Z.; Shehab A.; Tong Y.; Incremental Boosting Convolutional Neural Network for Facial Action Unit Recognition. *arXiv* **2017**. <https://arxiv.org/abs/1707.05395>
- [71] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision, *arXiv* **2015**. <https://arxiv.org/abs/1512.00567>
- [72] Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks *arXiv* **2018**. <https://arxiv.org/abs/1801.04381>
- [73] Radosavovic, I.; Kosaraju, R.P.; Girshick, R.; He, K.; Dollár, P. Designing Network Design Spaces *arXiv* **2020**. <https://arxiv.org/abs/2003.13678>

- [74] Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions *arXiv* **2017**. <https://arxiv.org/abs/1610.02357>
- [75] Katona, T.; Antal, B. Automated analysis of radiology images using Convolutional Neural Networks. In Proceedings of the 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA 2019), Dubrovnik, Croatia, 23–25 September 2019; pp. 89–92. <https://doi.org/10.1109/ISPA.2019.8868764>.
- [76] Katona, T.; Tóth, G.; Petró, M.; Harangi, B. Developing New Fully Connected Layers for Convolutional Neural Networks with Hyperparameter Optimization for Improved Multi-Label Image Classification. *Mathematics* **2024**, *12*, 806. <https://doi.org/https://doi.org/10.3390/math12060806>.
- [77] Katona, T.; Tóth, G.; Petró, M.; Harangi, B. Advanced Multi-label Image Classification Techniques Using Ensemble Methods. *MAKE* **2024**, **Bírálat alatt.**

Függelék

CITI vizsga jelentés

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM) COMPLETION REPORT - PART 1 OF 2 COURSEWORK REQUIREMENTS*

* Scores on this [Requirements Report](#) (Part 1) reflect quiz completions at the time all requirements for the course were met. The Transcript Report (Part 2) lists more recent quiz scores, including those on optional (supplemental) course elements.

• **Name:** Tamas Katona (ID: 7905396)
• **Institution Affiliation:** Massachusetts Institute of Technology Affiliates (ID: 1912)
• **Institution Email:** katona.tamas@inf.unideb.hu
• **Institution Unit:** Faculty of Informatics

• **Curriculum Group:** Human Research
• **Course Learner Group:** Data or Specimens Only Research
• **Stage:** Stage 1 - Basic Course

• **Record ID:** 30571891
• **Completion Date:** 14-Feb-2019
• **Expiration Date:** 13-Feb-2022
• **Minimum Passing:** 90
• **Reported Score*:** 100

| REQUIRED AND ELECTIVE MODULES ONLY | DATE COMPLETED | SCORE |
|--------------------------------------------------------------------------------------------|----------------|------------|
| Belmont Report and Its Principles (ID: 1127) | 13-Feb-2019 | 3/3 (100%) |
| History and Ethics of Human Subjects Research (ID: 498) | 13-Feb-2019 | 5/5 (100%) |
| Basic Institutional Review Board (IRB) Regulations and Review Process (ID: 2) | 13-Feb-2019 | 5/5 (100%) |
| Records-Based Research (ID: 5) | 13-Feb-2019 | 3/3 (100%) |
| Genetic Research in Human Populations (ID: 6) | 13-Feb-2019 | 5/5 (100%) |
| Populations in Research Requiring Additional Considerations and/or Protections (ID: 16680) | 14-Feb-2019 | 5/5 (100%) |
| Research and HIPAA Privacy Protections (ID: 14) | 14-Feb-2019 | 5/5 (100%) |
| Conflicts of Interest in Human Subjects Research (ID: 17464) | 14-Feb-2019 | 5/5 (100%) |
| Massachusetts Institute of Technology (ID: 1290) | 14-Feb-2019 | No Quiz |

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

This document was generated on 14-Apr-2024. Verify at:
www.citiprogram.org/verify/7kdfa65747-4d92-45b9-8eee-79ebf921fb90-30571891

Collaborative Institutional Training Initiative (CITI Program)
101 NE 3rd Avenue
Suite 320
Fort Lauderdale, FL 33301 US

Email: support@citiprogram.org
Phone: 888-529-5929
Web: <https://www.citiprogram.org>

Összefoglaló

Ebben a disszertációban a többcímekes osztályozás problémakörét vizsgáltuk meg mellkasröntgen felvételeken. A mellkasröntgen felvételek elemzése kritikus jelentőségű a különböző tüdő, szív és sok más betegség diagnosztizálásában. A többcímekes osztályozás lehetővé teszi, hogy egy adott röntgenfelvételen több betegség jelenlétét is detektálja. Ez különösen fontos, mivel egyetlen felvételen egyszerre több betegség is előfordulhat. Az általunk vizsgált megközelítések a mélytanulási modellek hatékonyságára és pontosságára építenek, különös tekintettel a CNN-ek alkalmazásával.

A 3. fejezetben egy többcímekes osztályozót építünk az alapoktól. Ennek az osztályozónak az alapját konvolúciós rétegek képezik. Különböző mélységű hálózatokat vizsgálunk képesek-e ezt az összetett kép osztályozási problémát abszolválni. A javasolt modellek képesek a nulláról megtanulni az alapvető címkéket de sajnos a state-of-the-art megoldásokkal nem vették fel a versenyt, így a további fejezetben ismert architektúrákkal haladtunk tovább. A fejezetben tárgyalt eredményeket a [75] publikációban mutattuk be.

A 4. fejezetben, egyelemű osztályozót alkottunk transzfer tanulás és hiperparaméter optimalizálás segítségével. A transzfer tanulás lehetővé teszi, hogy egy előzetesen betanított modellt, amely nagy mennyiségű adatot használt, újra felhasználjunk egy másik, hasonló feladatra. Itt ismert

CNN gerinceket: a ResNet, VGG és DenseNet architektúrákat alkalmaztunk alapmodellként. Megvizsgáltuk képesek-e röntgenfelvételek többcímzés osztályozására, majd transzfer tanulás segítségével inicializáltuk a legígéretesebbet. Erre az előtanított hálózatra építettünk egy teljesen új több rétegből álló osztályozó részt. Ez a módszer összehasonlításként szolgált ChestX-ray14 adatkészleten és eredményesebbnek bizonyult mint más csoportok által publikált eredmények. A [76] publikációban mutattuk be a fejezet eredményeit.

A 5. fejezetben továbblépve az egyelemű hálózatokon, egy többelemű ensemble osztályozót fejlesztettünk ki, amely különböző architektúrájú CNN osztályozók kombinációjából áll. Az ensemble modell célja, hogy javítsa az egyes modellek pontosságát és robusztusságát azáltal, hogy azok döntéseit kombinálja.

Az általunk javasolt új kombinált veszteségfüggvény az egyes tag modellek veszteségeit kombinálja a hálózat teljes veszteségével. Az így létrejött új ensemble osztályozó nem csak a mások által javasolt módszereket, hanem az általunk használt eddigi módszereket is felülmúlta. A harmadik rész eredményeit a [77] publikáció tárgyalja, amely jelen dolgozat írásakor elbírálás alatt volt.

Az itt bemutatott megközelítések új lehetőségeket nyitnak a gépi tanulás és az orvosi képfeldolgozás területén. Az eredmények azt mutatják, hogy a fejlett mélytanulási technikák hatékonyan alkalmazhatók a komplex orvosi

feladatok megoldására, és hozzájárulhatnak a későbbiekben a pontosabb és megbízhatóbb diagnózisokhoz, javítva ezzel a betegellátás minőségét.

Publikációs jegyzék

Folyóirat cikkek

Katona, T., Tóth, G., Petró, M., & Harangi, B. (2024). Advanced Multi-label Image Classification Techniques Using Ensemble Methods. *Machine Learning and Knowledge Extraction*, *bírálat alatt*, beküldve: 2024. május 07.

Katona, T., Tóth, G., Petró, M., & Harangi, B. (2024). Developing New Fully Connected Layers for Convolutional Neural Networks with Hyperparameter Optimization for Improved Multi-Label Image Classification. *Mathematics*, 12(6), 806.

Konferencia cikkek

Katona, T., & Antal, B. (2019, September). Automated analysis of radiology images using convolutional neural networks. In 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA) (pp. 89-92). IEEE.

Egyéb közlemények (nem része a disszertációnak)

Rebenku, I., Bartha, F. A., **Katona, T.**, Zsebik, B., Antalffy, G., Takács, L., ... & Vereb, G. (2023). Taking molecular pathology to the next level: Whole slide multicolor confocal imaging with the Panoramic Confocal digital pathology scanner. *Cytometry Part A*, 103(3), 198-207.

Besenczi, R., **Katona, T.**, & Szilágyi, M. (2015, October). A fork implementation of the Police Edition of the OOCWC system. In 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 163-164). IEEE.